Abortion Reporting and Underreporting: Can Better Design Yield Better Data?

Suzanne Bell, MPH

A doctoral dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Department of Population, Family and Reproductive Health
Johns Hopkins University Bloomberg School of Public Health

Baltimore, Maryland
March 14th, 2018

Abstract

Background

Induced abortion is a ubiquitous yet elusive phenomenon. It is among the most common health experiences, yet despite investigators' best efforts, we know relatively little about the specifics of its occurrence in most low-resource settings. In India, abortion is broadly legal but poorly measured. In this dissertation, we examined aspects of survey design and question methodology that could impact women's willingness to report abortion on face-to-face surveys.

Data and Methods

We used PMA2020 data from Rajasthan, India, which is a cross-sectional survey that is representative of women age 15 to 49. In Aim 1, we estimated the prevalence of induced abortion overall and for subgroups using list experiment and direct questions. In Aim 2, we assessed failures in abortion reporting using response time paradata from direct and list experiment questions. And in Aim 3 we sought to determine whether interviewer-respondent familiarity and respondent's prior survey experience were associated with improved reporting of abortion via the direct abortion questions.

Results

Despite having a large sample size of reproductive age women, the list experiment estimate of lifetime experience of abortion was actually significantly *lower* than the

direct abortion estimate (1.8% versus 3.5%). Further investigation into the list experiment assumptions revealed evidence of violations. We did not find evidence that poor numeracy or poor cognitive ability (as measured by schooling) explains the list experiment's poor performance. However, we did identify a significant editing effect whereby women who reported an abortion on the direct questions took 11.6 (95% CI 7.2-16.0) seconds longer to respond to the treatment list compared to women who reported no abortion on the direct questions. Regarding respondent acquaintance with the interviewer or the respondent's prior participation in a PMA2020 survey, we find these aspects of familiarity were not statistically significantly associated with abortion reporting, adjusting for respondent, interviewer, and community characteristics.

Conclusions

This dissertation provides a thorough investigation of abortion reporting and underreporting on a face-to-face survey in Rajasthan, India. It is also among the most in-depth studies of this phenomenon and the social and cognitive processes involved. Many of the specific analyses constitute the first investigations exploring these relationships with regard to abortion. As such, this dissertation contributes significantly to the literature on survey based abortion reporting. Many challenges remain in terms of our interest in identifying survey design features that would increase the validity of abortion reporting on face-to-face surveys.

Dissertation Committee

Committee of Final Thesis Readers

David Bishai, MD PhD
Professor and Thesis Advisor
Department of Population, Family and Reproductive Health
Johns Hopkins Bloomberg School of Public Health

Anne E. Burke, MD MPH
Associate Professor and Dissertation Defense Committee Chair
Department of Gynecology and Obstetrics
Johns Hopkins School of Medicine

Caroline Moreau, MD MPH PhD
Associate Professor
Department of Population, Family and Reproductive Health
Johns Hopkins Bloomberg School of Public Health

Elizabeth Colantuoni, ScM PhD
Associate Scientist
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Alternate Committee Members

Amy O. Tsui, MA PhD
Professor
Department of Population, Family and Reproductive Health
Johns Hopkins Bloomberg School of Public Health

Janice Evans, PhD
Professor
Department of Biochemistry and Molecular Biology
Johns Hopkins Bloomberg School of Public Health

Andreea Creanga, PhD
Assistant Professor
Department of International Health
Johns Hopkins Bloomberg School of Public Health

Acknowledgements

My time at Johns Hopkins has been incredibly challenging and rewarding. I am grateful for the opportunity to have studied at such a distinguished academic institution and for the financial support that allowed me to do so. Thank you to the funders of the Robertson Endowment, the Fund in Recognition of Laurie Schwab Zabin, the Endowed Fellowship in Family Planning and Reproductive Health, and the Edward J. Dehne Award in Population Dynamics.

Many thanks to my dissertation committee, Dr. Anne Burke, Dr. Caroline Moreau, and Dr. Elizabeth Colantuoni, as well as the alternate committee members, Dr. Amy Tsui, Dr. Janice Evans, and Dr. Andreea Creanga. I am also thankful to other faculty who served as members of my proposal defense committees, including Dr. Stefan Baral, Dr. Amy Tsui, Dr. Saifuddin Ahmed, Dr. Stephane Helleringer, and Dr. Li Liu. Their constructive comments and insights strengthened my proposal and subsequent dissertation. I would also like to thank the many Johns Hopkins faculty and staff who provided instruction and mentorship to me throughout the program, in particular Beth Fredrick.

I would like to thank my academic advisor, Dr. David Bishai, for his encouragement and confidence in my abilities. He has played an integral role in my development as an independent researcher and public health professional. I am incredibly grateful for all I have learned and will continue to learn from him.

My dissertation would not have been possible without the efforts of many individuals at PMA2020. Many thanks to Dr. Scott Radloff for his support, and to Dr. Anoop Khanna and the entire team at the Indian Institute of Health Management and Research for their hard work and dedication. I am also grateful to the many Rajasthani respondents for their time.

I would like to thank my friends and colleagues at Johns Hopkins and beyond who have provided helpful feedback, insight, and encouragement throughout the PhD program. These wonderful individuals include Hannah Olson, Laura Hinson, Susannah Gibbs, Meghan Gallagher, Abba Greenleaf, Sally Safi, Claire Greene, Bethany Warren, Blair Berger, Andrea Rowan, Elizabeth Flock, and Matt Dobson. I also want to thank my brilliant friends from UC Berkeley: Sarah Treves-Kagan, Mridula Shankar, Alison Gemmill, Lauren Lessard, and Karen Weidert. They have provided me with emotional and intellectual support throughout the years.

I would especially like to thank my husband for his patience, his cooking, his excellent editing skills, and his endless love and support during this time and in life generally; he is my favorite. I am also thankful to Chuck Belldig, my furriest friend, who has been a steadfast companion and reluctant snuggler during the dissertation writing process.

Finally, I am grateful for my family and each of the incredibly thoughtful, funny, and intelligent individuals in it. In particular, I want to thank my amazing "little" brother, who can always make me laugh. And most importantly, I would like to express profound gratitude to my parents for providing me with a foundation of unconditional love, for fostering in me an inquisitive mind, and for demonstrating how to approach life with strength, compassion, a sense of humor, and a melody. I love them dearly.

Table of Contents

List of Tables and Figures

1. Introduction

1.1 Background

A fundamental problem in abortion research is accurately measuring its occurrence and the characteristics of the women who experience it. The ability to accurately quantify pregnancy rates, unintended pregnancy rates, and contraceptive method failure rates relies on the capacity to measure the incidence and descriptive demography of induced abortion within a specific geography and time period (Jones and Kost 2007). The responsiveness of evidence-based policies and programs to improve reproductive health services also relies on understanding abortion utilization. Abortion data in low-resource settings are often estimated through personal interviews in population-based sample surveys (Guttmacher Institute 2010). Social stigma around the practice of induced abortion often leads to under- and misreporting of terminations, even where the procedure is legal (Guttmacher Institute 2010; Jones and Kost 2007). Scientific assessments of survey methods' ability to measure sensitive behaviors have focused on the extent of truthfulness in responses (Guttmacher Institute 2010).

Researchers have long sought to determine whether modifying aspects of survey design or employing different question methodologies improves reporting of abortion and other sensitive items (Elul 2004; Grossman et al. 2015a; Guttmacher Institute 2010; Lensvelt-Mulders et al. 2005; Rossier 2003; Rossier et al. 2006; Tourangeau and Yan 2007; Yeatman and Trinitapoli 2011). Related work has advanced our understanding of the potential mechanisms of underreporting in survey research involving topics that invoke socially desirable responses (Holden et al. 1992; Holtgraves 2004; Holtgraves et al. 1997; Sudman et al. 1996). One innovative survey methodology is the list experiment, also known as item count technique, which offers a statistical means to measure sensitive items through indirect questioning (Glynn 2010). The list experiment is structured to reduce the social desirability pressures of direct questioning and underreporting of stigmatized behaviors (Blair and Imai 2012; Glynn 2010; Imai 2011; Moseson et al. 2015). Respondents receive either the control or treatment version of a list of items that are identical except that the treatment version includes the sensitive item, for example induced abortion. Respondents are asked to report *how many*, not *which* items they have ever experienced (prevalence) or experienced in the previous year (incidence). In seeking to improve abortion reporting through the list experiment, we can also benefit from greater understanding of the social and cognitive psychology behind respondents' experience answering sensitive survey questions.

Another aspect of the survey design is the potential to use enumerators from the communities surveyed. Interviewer-respondent familiarity and the respondent's experience with prior surveys may have an impact on respondents' willingness to report stigmatized behaviors, like abortion (Rodriguez et al. 2015; Safi et al. 2017; Sana et al. 2016; Stecklov et al. 2015). Limited empirical evidence finds that respondent familiarity with the interviewer may actually improve reporting of sensitive items (Rodriguez et al. 2015; Safi et al. 2017; Sana et al. 2016; Stecklov et al. 2015).

This dissertation research seeks to assess the performance of the list experiment for measuring induced abortion in Rajasthan, India. This investigation will be the first to assess the feasibility and acceptability of using the list experiment to collect self-reported data on induced abortion incidence, and will be the first to more thoroughly investigate the performance of the list experiment to measure induced abortion prevalence in comparison to direct estimates of abortion from the same population. With these data, we aim to generate new insight into the social epidemiology of abortion in a low-resource setting as well as the social and cognitive psychology involved in answering survey questions on sensitive topics.

We added the direct and list experiment abortion questions to a representative state-level household survey of some 6,000 females of childbearing age in Rajasthan, India. Elective termination of pregnancy is legal in India but official Rajasthan government statistics show only 2 per 1,000 women age 15 to 49 experience an induced abortion annually, based on data from certified facilities (Ministry of Health and Family Welfare 2013). Because a substantial number of inductions are performed in private facilities or are self-induced, many abortions are not captured in these statistics (Elul et al. 2004). Population-based surveys have the potential to capture abortions missed by official statistics; however, estimates from direct questioning can be affected by social desirability bias. In Rajasthan, only 3% of pregnancies end in abortion (spontaneous or induced) according to population-based survey results (Office of the Registrar General & Census Commissioner et al. 2013). These estimates likely suffer from significant underreporting due to the stigma associated with the procedure. Differential underreporting by vulnerable women would make survey estimates difficult to recalibrate for reporting bias (Duggal and Ramachandran 2004; Jones and Kost 2007).

1.2 Specific Aims

The overall objective of this dissertation is to assess induced abortion reporting and underreporting on a population-based survey of reproductive age Rajasthani women. The current dissertation research has three specific aims.

Aim 1: Estimate the prevalence of induced abortion overall and for subgroups using list experiment and direct questions. These estimates enable comparison of each methodology's performance overall and by subgroup.

Aim 2: Assess failures in abortion reporting using response time paradata from direct and list experiment questions. This aim leverages the paradata (i.e. survey process data) available via the smart phone time stamped log, allowing us to examine the underlying mental processes involved in responding to cognitively demanding survey questions.

Aim 3: Determine whether interviewer-respondent familiarity and respondent's prior survey experience are associated with improved reporting of abortion via direct abortion questions. The study uses local interviewers to implement the survey, thus interviewers are acquainted with many respondents. Additionally, the cross-sectional survey is administered repeatedly in communities every 6 to 12 months, thus some respondents are randomly selected for participation more than once. We aim to explore the potential impact of these survey design features on abortion reporting.

1.3 Overview

The second chapter of this dissertation presents a review of the existing literature on sources of bias in reporting of sensitive behaviors on surveys and the various methods used to measure induced abortion, with a focus on the list experiment. Given the unique structure of list experiment questions, we also present social and cognitive psychology literature on the stages of responding to survey questions and discuss the specific application of this paradigm to the list experiment questions. The third chapter includes a description of the data source, a Performance, Monitoring and Accountability 2020 (PMA2020) survey conducted in Rajasthan, India, and the associated survey design, sampling, piloting, and training details. The next three chapters present results from the analyses corresponding to the aforementioned three specific aims. The final chapter is a discussion of the findings, synthesizing the results with regard to the broader public health and policy context.

2. Background

2.1 Significance

Pregnancy termination occurs in all societies and predates the modern era, but its occurrence is often clandestine and thus difficult to measure (Guttmacher Institute 2010; Joffe 2009). Although a precise count is unknown, recent estimates indicate that approximately 56.3 (90% uncertainty interval [UI] 52.4 to 70.0) million induced abortions occurred annually in 2010 through 2014, and researchers estimate the global annual incidence of induced abortion is 35 (90% UI 33 to 44) per 1,000 women age 15 to 49 (Sedgh et al. 2016). The induced abortion rate in the developed world declined from 46 (90% UI 41 to 59) to 27 (90% UI 24 to 37) between 1990-1994 and 2010-2014, but researchers found a non-significant 2 point decline from 39 (90% UI 37 to 47) to 37 (90% UI 34 to 46) during the same time period in developing countries (Sedgh et al. 2016). The vast majority of abortions in low- and middle-income countries (LMIC) continue to be unsafe (Ganatra et al. 2017).

Although induced abortion is common, often only regional estimates are available (e.g. Western Africa, South America, etc.) and are generated from a mix of data sources (of varying quality), assumptions, and imputation models (Sedgh et al. 2016; Sedgh et al. 2012). Investigators have used many survey-based approaches to try to measure induced abortion but none have effectively resolved all of the many impediments to eliciting accurate responses. Twenty-six percent of the world's population currently resides in a country where abortion is generally prohibited

(Center for Reproductive Rights 2013). In these countries, and even in many countries where abortion is largely legal, reliable vital statistics or surveillance data on induced abortion are deficient or altogether absent (Stillman et al. 2014). As a result, these countries have limited knowledge about the extent of induced abortion locally, trends in termination rates, the demography of women who terminate pregnancy, abortion safety, and risk-factors for abortion-related morbidity and mortality as a result of unsafe abortion.

The ability to design evidence-based policies to improve family planning services and safe abortion or post-abortion care, as well as to normalize the procedure, depends on the capacity to measure induced abortion within a specific geography and time period. Abortion incidence measurement studies conducted by the Guttmacher Institute have contributed to fuller implementation of safe abortion services in Ethiopia, and have helped lay the foundation for abortion law reform in Columbia, Mexico City, and Rwanda (Singh and Wittenberg 2017). Additionally, abortion prevalence and incidence data enable improved estimation of unwanted pregnancy, which is a key measure of the quality of reproductive health services and the impact of family planning programs and policies (Singh et al. 2010).

Beyond measurement of the behavior itself, information about the social and cognitive experience of being asked about one's past experience with abortion on a survey can help researchers better understand how and why the phenomenon is underreported by women. These findings can inform future survey design in an effort to improve induced abortion reporting in related data collection activities.

2.2 Sources of Bias in Reporting

Sensitive items in surveys are those that respondents perceive to be intrusive, that raise concerns about the threat of disclosure and associated possible consequences, and that could elicit socially undesirable responses (Tourangeau and Yan 2007). Asking sensitive questions can affect the validity of survey responses in three ways: 1) the overall response rate, i.e. the percentage of respondents who agree to participate in the survey; 2) item non-response rate, i.e. the percentage of respondents who do not respond to a given question but who agree to participate in the survey; and 3) response validity, i.e. the percentage of respondents who respond truthfully to the sensitive question(s) (Tourangeau and Yan 2007). In nearly all societies, people consider survey questions on induced abortion sensitive. As such, results from questions on induced abortion can be biased to varying degrees through all three of the aforementioned mechanisms if these reporting issues occur in a systematic way. The first two sources of bias can be investigated empirically rather easily. High overall non-response and item non-response may be indicative of respondent reluctance to answer questions concerning sensitive topics, which could introduce bias. The third potential source of bias in the reporting of sensitive behaviors is more difficult to assess and has been the motivation for much research investigating the reason(s) for sensitive behavior underreporting. This research has largely focused on the extent to which people respond to sensitive items in a socially desirable manner.

Social desirability can affect respondents' answers regarding sensitive topics through different mechanisms, but it is important to first describe survey interactions and the process by which individuals answer survey questions. The survey is a social phenomenon that involves cognitive exertion on the part of the individuals involved and is ultimately governed by the same linguistic and social norms as a conversation (Sudman et al. 1996). The four principles that guide conversations thus also guide the survey interaction (Grice 1975). These principles are as follows: 1) speakers should be honest; 2) speakers should make comments that are relevant to the conversation; 3) speakers should make informative and non-repetitive contributions to the conversation; and 4) speakers should be clear (Grice 1975). These tacit expectations inform the responses individuals provide to questions by affecting multiple aspects of the cognitive process involved in answering survey questions.

Although it can take mere seconds to provide an answer to a survey question, responding to self-report behavior items involves a number of stages generally thought to include: 1) interpreting the question; 2) retrieving the information (and generating a representation of the behavior); 3) formatting a response; and 4) editing a response (Figure 2.1) (Sudman et al. 1996).

Figure 2.1. Stages of responding to self-report behavior survey questions

Source: Sudman et al. 1996

Interpreting the question requires an understanding of both the literal and the pragmatic meaning of the question. Respondents identify words and phrases and construct a meaning within the context of the survey. Questions can involve ambiguous words or phrases and complex structures that the respondent must interpret and make a decision regarding their interpretation. This process can be affected by antecedent questions and other aspects of the survey and is generally governed by the conversation assumption that the interviewer asks questions to get new information.

Once the respondent has interpreted the question, retrieving information for their answer can be a demanding task. Respondents are ultimately "cognitive misers" who tend to follow a "satisficing" rather than "optimizing" strategy (Krosnick 1991; Taylor 1981). The ability to retrieve the necessary behavioral information can be aided by cuing the topic in preceding questions but generally it is difficult to induce respondents to engage more thoroughly in the retrieval process if they are not willing to exert the effort required (Sudman et al. 1996). Forgetting information can occur if the mind did not pay attention to or encode the event initially, if the memory has not been recalled for a long time thus making it difficult or impossible to retrieve even if it still exists in one's memory, or the memory can be altered or erased between when it occurred and subsequent recall and re-storage events (Sudman et al. 1996).

Formatting the response is the process by which respondents take the information they retrieved and translate it into an answer. At this stage, one may factor in the response options (if it is a close ended question), the previously determined meaning and intent of the question, and the information that has already been provided on previous survey questions in forming their potential response. In the final stage, individuals may alter their response before sharing it with the interviewer. This editing process reflects the presence of social desirability and self-presentation (Holtgraves 2004; Nowakowska 1970), but social desirability can also affect the other stages previously described.

Researchers have sought to measure the impact of respondents' social desirability. Some view social desirability as a stable personality trait and others view it as a temporary social strategy in the context of a survey; survey researchers tend to view it as the latter (Tourangeau and Yan 2007). Attempts to measure social desirability are difficult because it is challenging, if not impossible, to distinguish between respondents who are truly compliant with the social norm, respondents who have a sincere but exaggerated or incorrect perception of themselves (or their past behaviors) that results in a faulty retrieval process, and respondents who are deliberately seeking to maintain a favorable impression by inaccurately responding to items (Tourangeau and Yan 2007). The latter two possibilities, termed by Paulhus

(1984) as self-deception and impression management, respectively, are most often identified as the underlying measures of socially desirable responding in empirical analyses, such as factor analyses (Paulhus 1984). Researchers can detect self-deception indirectly via response times. Individuals whose responses coincide with a certain false image of themselves take less time to respond (Holden et al. 1992). Similarly, respondents who don't want to consider themselves as having a socially undesirable trait may be unwilling to engage in the retrieval process and take less time to respond (Holden et al. 1992). Thus, it takes less time to lie than to retrieve.

Alternatively, researchers can assess social desirability operating through impression management by determining if respondents take *longer* to respond to the sensitive questions than similar non-sensitive items. The longer response time is thought to indicate that the sensitive question triggers a more controlled response process whereby respondents edit their answers in a socially desirable manner. Empirical research supports this idea of a deliberate editing process, with evidence indicating that increasing sensitivity or social desirability concerns is associated with increasing response time (Holtgraves 2004; Holtgraves et al. 1997). Despite these indirect means of assessing the three main sources of bias in reporting sensitive items, other factors could also be responsible for these findings, thus it is not absolute that higher non-response and longer response times lead to bias.

2.3 Sensitive Questions in Surveys

The principal concern when asking sensitive questions in surveys is the accuracy of the responses. Accuracy is comprised of validity and reliability. Internal validity is the extent to which a study is free from bias or systematic error. This type of validity is a prerequisite for external validity, which is the degree to which the results of a study can be generalized to populations or groups that did not participate in the study. Random error in surveys reduces the reliability or precision of the results but does not introduce bias. Data must be both valid and reliable in order to be accurate.

Answers to sensitive questions on surveys often suffer from significant underreporting. Resulting population-based prevalence or incidence estimates are thus not internally *or* externally valid. However, if underreporting is random, the observed relationships between the sensitive behavior and other respondent characteristics could still be valid. Only if underreporting is *differential* depending on respondent characteristics are the resulting data biased. Thus, a desirable survey design or question methodology is one that produces the most reliable and valid prevalence or incidence estimates, as well as the most valid estimates of relationships between the sensitive behavior of interest and other variables. Because the main concern with survey questions on sensitive topics is underreporting, the most valid estimate is often the highest estimate. However, this may not always be the case.

In most instances, there is no gold standard measure against which researchers can assess the validity of a survey based sensitive behavior estimate. When this is the case, investigators often rely on convergent validity, which is the degree to which two measures that theoretically should be related are in fact related. Convergent validity is a component of construct validity, or the degree to which a construct measures what it purports to be measuring. Conversely, researchers can use discriminant validity, which determines whether concepts that should *not* be related are actually unrelated. In measuring lifetime prevalence of sensitive behaviors, one way to assess convergent validity is to determine whether the prevalence uniformly increases with age.

Many aspects of survey design affect reporting on sensitive topics. The primary factors affecting disclosure include mode of survey administration, the question methodology, and the data collection setting (including whether the interviewer or others are present). The most common methods of survey administration are face-to-face interviews, phone interviews, and mail interviews. Internet surveys are also becoming increasingly common. One key distinction across mode of administration is whether an interviewer administers the questionnaire. Studies have repeatedly demonstrated that respondents are more willing to report sensitive items in the context of self-administered questionnaires (SAQ) as opposed to interviewer administered (Tourangeau and Yan 2007); this has been established in the case of induced abortion reporting specifically (Lensvelt-Mulders et al. 2005). With regard to self-administration survey type, there appears to be no statistically significant improvement in reporting sensitive items when the survey is conducted via computerization as opposed to a paper questionnaire, although computerization does tend to result in slightly higher estimates (Tourangeau and Yan 2007). In low-resource and low-literacy populations though, typically interviewer administered, face-to-face surveys are conducted.

Related to the mode of survey administration is the method of asking about the sensitive item. Direct questions, regardless of survey mode, are known to result in substantial underreporting of sensitive items (Jones and Kost 2007; Tourangeau and Yan 2007). Thus, indirect methods for eliciting responses to sensitive items are a strategy used to reduce the incentive to answer questions in a socially desirable manner.

The Anonymous Third-Party Reporting (ATPR) methodology asks respondents about their confidants' experiences with the sensitive item of interest. Results indicate the utility of this method is context dependent, performing better in settings where it is normative that people share details about the sensitive item with their network of close friends or relatives (Elul 2004; Rossier et al. 2006). The best-friend method, a version of the ATPR that only asks about one best friend's experience with the sensitive item, appears to more consistently outperform direct questioning results (Grossman et al. 2015b; Yeatman and Trinitapoli 2011).

Randomized response technique (RRT) is another indirect survey methodology. Using RRT, respondents conduct the randomization process and, depending on the outcome, either respond "yes" or "no" to a question with known probability (e.g. Were you born in April?) or "yes" or "no" to the sensitive item (e.g. Have you ever had an abortion?). The critical aspect of this approach is that the interviewer does not know which question the respondent is answering. Meta-analysis results indicate that RRT is effective in reducing underreporting of sensitive behaviors compared to direct questioning — this is true in validation studies as well as comparisons to other methodologies (Lensvelt-Mulders et al. 2005).

Similar to RRT, endorsement experiments offer an indirect way of asking sensitive questions on surveys by obscuring individual responses. It draws on extensive psychological research on persuasion and evaluation bias, which indicates that people are more likely to be influenced by likeable sources and people's endorsements are much more effective when an individual has a positive attitude towards the endorser (Blair et al. 2014; Rosenfeld et al. 2014). Respondents are divided into two groups; the control group is asked to rate their level of support for a non-sensitive item whereas the treatment group is asked the same question except the non-sensitive item is said to be associated with the sensitive item. The difference between the two groups is then taken to approximate the degree to which respondents are favorable (or unfavorable) towards the sensitive item. Evidence around the endorsement experiment remains mixed; it has been shown to perform similarly to other indirect survey methodologies in some instances but better in others (Blair et al. 2014; Rosenfeld et al. 2014).

In addition to the mode of administration and question methodology, one can modify the interview context and other survey strategies to improve reporting of sensitive items. Conducting interviews in a private setting where the respondent cannot be overheard is standard practice so as to avoid potentially negative effects of others' presence on data quality. Studies have found that parental presence significantly reduces children's reporting of socially undesirable responses, although spouse presence has no significant effect (Tourangeau and Yan 2007). Wording questions in a "forgiving" manner may increase reporting of socially undesirable behaviors, although few studies have validated use of this method. This tactic uses non-judgmental language in a preamble to the sensitive question; for example, a question about voting may be prefaced with, "We often find a lot of people were not able to vote because they weren't registered, they were sick, or they just didn't have the time." Other approaches that appear to improve reporting accuracy include matching interviewers and respondents on characteristics like sex, race, and ethnicity (Catania et al. 1996; Davis et al. 2010; Sudman and Bradburn 1974; Weinreb 2006), assurances of data confidentiality (Singer 1995), priming the respondent to answer honestly (Cannell et al. 1981), and putting the sensitive questions towards the end of the survey to maximize interviewer-respondent rapport (Fowler Jr 2013).

Specifically with regard to interviewer and respondent familiarity, the limited empirical evidence exploring this topic indicates that prior acquaintance does not reduce data quality and may actually improve it. Researchers conducting longitudinal data collection in Kenya found that interviewers who knew the respondent well or very well produced more consistent data across survey waves than strangers (Weinreb 2006). In data from the Dominican Republic, familiarity had no effect on response patterns for most variables, but it did reduce non-response and improve validity for some (Rodriguez et al. 2015; Sana et al. 2016). Also in the Dominican Republic, researchers found that local interviewers, both those who knew the respondent and those who did not, collected more realistic female sterilization data than interviewers from outside the community (Stecklov et al. 2015). More recent results using data from a multi-country study that employs resident enumerators to conduct repeated cross-sectional family planning surveys found that prior acquaintance and participation in a previous cross-sectional survey does not appear to negatively impact data quality and may improve it (Safi et al. 2017).

Ultimately there is no generally agreed upon standard method for eliciting honest responses to sensitive items on surveys, but rather a number of factors to consider depending on the respondent population, type of information desired, and funds available.

2.4 Current Induced Abortion Measurement Approaches

Specifically related to induced abortion measurement, we know there is substantial underreporting from direct questioning. In particular, there is differential underreporting of induced abortion by certain women, thus demographers cannot easily calibrate survey estimates to account for this downward bias (Jones and Kost 2007). Researchers have applied or developed several survey-based approaches to the measurement of induced abortion but none have successfully addressed all of the many obstacles to eliciting accurate reporting (Guttmacher Institute 2010; Huntington et al. 1996; Rossier 2003). Even in settings where abortion is legal and less stigmatized, the proportion of known abortions that direct questioning in face-to-face interviews captures ranges widely from only 35% to 48% in the United States, to 73% in Estonia (Anderson et al. 1994; Jones and Kost 2007).

The use of SAQ and audio-SAQ have reduced the level of underreporting, but may result in questionable data quality as a result of poor understanding and incorrect utilization of skip patterns (Jones and Forrest 1992; Langhaug et al. 2011). Audio computer-assisted self-interview (ACASI) typically results in higher reporting of sensitive sexual behaviors than face-to-face interviews (although not always) and researchers can utilize this method in low-literacy populations (Langhaug et al. 2011; Lara et al. 2004). In Madhya Pradesh, India, investigators used a mixed methods narrative approach that generated an abortion ratio five times higher than that obtained from direct questioning (Edmeades et al. 2010). This technique is very costly and time intensive and has not been tested elsewhere.

Empirical evidence indicates that RRT is the survey approach that most often results in the highest proportion of respondents reporting an induced abortion or other sensitive behavior (Coutts and Jann 2011; Lara et al. 2004). The drawback of this approach is that respondents report lower trust in the confidentiality of their response and this technique is difficult and time consuming to train on and implement

(Coutts and Jann 2011). The ATPR has had mixed results depending on the context in which researchers have implemented it. In India, the ATPR generated suspect estimates (Elul 2004; Rossier et al. 2006). The limited available evidence regarding the best-friend method suggests that it provides improved estimates of induced abortion, but further research is needed (Grossman et al. 2015b; Yeatman and Trinitapoli 2011). The choice of which method to use when estimating induced abortion is context dependent and there is currently no gold standard in the field (Rossier 2003).

2.5 Dissertation Research

In a list experiment, investigators can overcome many of the challenges faced by other survey-based approaches. Individuals can know that their responses will be confidential, reducing the social desirability pressures of direct questioning, and this is achieved simply, with limited additional training or cost if embedded in an existing survey. Also, there now exist multivariable analysis options for the list experiment, which earlier had not been the case (Blair and Imai 2012; Corstange 2008; Glynn 2013; Imai 2011). Researchers have increasingly used list experiments in place of RRT given that list experiments can be easier to implement and understand (Coutts and Jann 2011). Additionally, studies have shown that respondents trust and accept questions in the list experiment format more so than RRT (Coutts and Jann 2011). Results produced by the list experiment also have lower item non-response and can be more reliable than results from RRT, particularly for highly sensitive behaviors (Coutts and Jann 2011).

The current dissertation research included list experiment questions on a representative survey of reproductive age women in Rajasthan, India. There is growing use of this method in studies of abortion. But a list experiment has only been applied to the measurement of induced abortion prevalence in a representative sample on one prior occasion, and it has never before been used to estimate induced abortion incidence. In Liberia, researchers recently used a list experiment to measure lifetime experience of abortion nationally and produced an abortion prevalence of 35% (Moseson et al. 2015). However, the study omitted in-depth exploration of violations of list experiment assumptions and there was no contemporaneously collected direct abortion estimate for comparison (Moseson et al. 2015). More recently, researchers conducted an online pilot employing the list experiment for measurement of lifetime experience of abortion on a convenience sample of 1,233 adult women (Cowan et al. 2016). The resulting list experiment abortion estimate was 22% while the direct question abortion estimate was 18% (Cowan et al. 2016). These researchers also did not conduct an in-depth exploration of list experiment assumptions. More recent applications of the list experiment to measure abortion in Vietnam and Texas had less positive results, and the recommendations for further research were more measured (Moseson et al. 2017b; Treleaven et al. 2017).

The current dissertation aims to improve upon this previous research and more rigorously evaluate the use of a list experiment in the measurement of induced abortion. The results will provide much needed evidence as to the quality of the data collected from this technique, types of biases that are present, and insights into potential statistical adjustments. Better measurement of the demography, prevalence, and incidence of abortion can shed light on how this behavior fits in to women's repertoire of pregnancy management choices and bring attention to inequities in access to family planning and safe abortion services.

2.6 The List Experiment

The list experiment methods were originally developed in the 1980s by social psychologists to elicit truthful responses to sensitive questions (Kuklinski et al. 1997; Miller 1984; Sniderman and Carmines 1997). The standard list experiment randomizes individuals to either the treatment or control group. The control group is read a list of non-sensitive items, whereas the treatment group is read the same list, plus the sensitive item. Respondents are then asked to report *how many* of the items they have ever experienced (prevalence), not *which* ones, without directly mentioning each item. A simple difference in means between the mean total item counts of the treatment and control groups is then calculated. The double list experiment is a modification whereby every respondent receives a treatment version of one list and a control version of another list, thus everyone serves as control and treatment within the sample. This is a more efficient estimator than the standard list experiment.

Most of the literature assessing this methodology compares list experiment estimates to those obtained via direct questioning or other survey methodologies. The general "more is better" assumption that higher estimates are more valid is typically used to assess performance (Tourangeau and Yan 2007). Results from several empirical studies illustrate that the list experiment significantly outperforms direct questioning across multiple modes of administration, particularly with more sensitive behaviors (Aronow et al. 2015; Comşa and Postelnicu 2013; Gonzalez-Ocantos et al. 2012; Tsuchiya et al. 2007; Wolter and Laier 2014). Specifically in the two face-to-face surveys that employed the list experiment (both of which were in low-resource settings), list experiment estimates were higher than those obtained via direct questioning (Wolter and Laier 2014). In addition, interviewers reported greater comfort in asking the list experiment questions than the direct questions (Wolter and Laier 2014).

Despite these findings, which support use of a list experiment, there are several studies where list experiments failed to produce more valid estimates of the sensitive behavior(s) (Biemer and Brown 2005; Coutts and Jann 2011; Droitcour et al. 2004; Rosenfeld et al. 2014). A recent study using population level (as opposed to individual level) objective measures of the sensitive behavior found that, although list experiment estimates resulted in estimates that were closer to the truth than those obtained from direct questioning, RRT and endorsement experiment estimates were even more valid (although the endorsement experiment confidence intervals were largest) (Rosenfeld et al. 2014). Further validation studies measuring a range of sensitive behaviors in different populations are needed to better understand when the list experiment is likely to outperform or underperform other methods.

2.7 Responding to a List Experiment

Drawing on the aforementioned social and cognitive psychology theories regarding how respondents answer survey questions, we have outlined a theory specific to how respondents answer list experiment questions. We posit that answering a list experiment question involves the following stages: 1) decode the instructions; 2) interpret the question; 3) silently retrieve and recall past life events from memory to consciousness; 4) form a response by enumerating the events; and 5) make a strategy to edit one's response or not (Sudman et al. 1996). Failures can occur at any of the five stages when answering the list experiment questions and these failures may be more likely given the particularly stigmatizing and at times ambiguous nature of abortion.

In stage one, the more complex instructions associated with the list questions may result in respondents having difficulty understanding the task at hand. We anticipate women with less education would have more difficulty with the instructions. We also anticipate that women with less education who have difficulty with the instructions would require that the instructions be re-read and will thus take more time on the instructions related comprehension question. They would also take longer to work through the example list experiment question that involves foods eaten in the previous week.

Once the respondent understands the instructions and has answered the example question, she must interpret each of the actual list experiment questions. This involves understanding each item on a given list, although we are primarily concerned with the interpretation of the sensitive item of interest, i.e. whether the respondent has had an induced abortion. Due to the randomization, people receiving the control and treatment versions of a given list will interpret or misinterpret the control items similarly, on average.

It may be especially difficult to include abortion as a list experiment item given what we know about the myriad ways people may conceive of or interpret an experience that a researcher would consider an induced abortion. Some women use specific local words or phrases to refer to the abortion experience, but would not call it an abortion. For instance in Tanzania, women may use traditional medicines to “move a pregnancy to the back”, thus suspending it indefinitely until a time when the woman wishes to continue the pregnancy (Plummer et al. 2008). There is also evidence to suggest that some women may genuinely conceive of medical abortion as a miscarriage even though they took medicine to intentionally bring about the miscarriage-like experience (Kanstrup et al. 2017; Simonds et al. 1998). And in some contexts, women take pills or have a procedure to “bring back their period” (or “wash”) without confirming whether they are pregnant, referred to as “menstrual regulation” in many settings; this is not considered an abortion although in practice it is (Rahman et al. 2014). The lack of pregnancy confirmation and corresponding data on this aspect of menstrual regulation makes it somewhat unclear even to the researcher how one should categorize this event, although it is generally believed that women correctly presume they are pregnant (Rahman et al. 2014). Thus, understanding the list experiment question goes beyond simply understanding the words used by the researcher; the woman must in particular understand the broader experiences meant to be encapsulated by the "had an induced abortion" item. Failures at this stage can easily occur given the many ways that women conceive of and talk about abortion, which can vary within and across countries as well as over time in the same context.

After interpreting the questions and list items, retrieving and recalling the past life events included on a list experiment list is the third stage of the response process. We are most concerned about the retrieval of the "had an abortion" item, thus we focus on the potential failures that can occur with regard to its retrieval. As previously mentioned, women may think about an abortion in many ways at the time of the experience, and this notion may change over time. To the extent that women do not initially conceive of an experience as an abortion and thus do not mentally code it as such, women would "correctly" not retrieve and recall that experience in the count of items they have experienced (Kanstrup et al. 2017; Simonds et al. 1998). Alternatively, women may initially consider an experience an abortion but that conceptualization changes as time passes. There is evidence to suggest that women who had a medical abortion may be more likely to subsequently report it as a miscarriage than if they had a surgical abortion (Foster 2016). Lastly, women may fail to retrieve a relevant memory. This can occur if a woman genuinely forgot the experience, or may be a result of social desirability whereby a woman does not even engage in the retrieval process because she would not admit to it even if she had experienced this socially stigmatized behavior. If the latter occurs, evidence suggests that this faulty or biased retrieval may reveal itself via shorter response times (Holden et al. 1992).

Following the retrieval and recall process, women must enumerate the number of items they determined having experienced. This requires some level of numeracy and is cognitively demanding. As such, failures can occur here due to limited quantitative skills or limited effort (Krosnick 1991). In this event, a woman may indicate to the RE which specific events she has experienced, or she may simply provide an estimate of the number of items she has experienced.

Lastly, after a respondent determines the number of items she has ever experienced, she may decide to edit her response factoring in the response options and social desirability pressure. For instance, if her number is the highest potential response, she may decide to lower the number so as not to reveal that she has in fact had an abortion. Alternatively, she may have experienced 0 items but think this reveals a lack of interaction with the healthcare system. This could also be deemed stigmatizing, causing her to edit her numeric response upward. Other mental calculations and strategizing may take place in this last stage that causes the respondent to report a number other than the one she determined in the previous step. Evidence suggests that this deliberate editing process can present itself in the form of longer response times (Holtgraves 2004).

Below is a table that summarizes the aforementioned forensic patterns left by failures at each step, along with the means of detecting these failures (Table 2.1).

Table 2.1. Stages of list experiment responding and evidence of failures

Stage	Failure pattern	Null hypothesis
1) Decode the instructions	Groups with more schooling take less time to answer question regarding whether understand list experiment instructions	Equality of time spent listening to list experiment instructions and confirming understanding by schooling
	Groups with more schooling will take less time to answer the example list experiment question	Equality of time spent on list experiment example question by schooling
2) Interpret the question	No clear pattern; would need to conduct "think aloud" qualitative cognitive interview to determine how respondent interpreted the list experiment question	N/A
3) Retrieve and recall past life events	May intentionally skip or truncate this stage and simply estimate list experiment response, representing a "fake" or biased retrieval process	Equality of response time for people who express more social desirability pressure (not readily detectable from available data)
4) Enumerate the events	Groups who respond quickly on other questions that require similarly high levels of numeracy and cognitive ability will respond quickly to the list questions	No difference between time spent on similarly cognitively demanding questions and time spent on list experiment questions
5) Make a strategy regarding a response	Respondents who report an abortion on the direct question take longer to finalize their list experiment response if	Equality of time spent on treatment list experiment questions by direct abortion response

	they considered editing their answer	

2.8 Specific Aims

The current dissertation work draws on existing survey methodology literature and social and cognitive psychology in an attempt to improve self-report of induced abortion. We begin by conducting a comparison of abortion reporting utilizing direct and list experiment questions and explore potential explanations for failures of the list experiment assumptions. We then use survey paradata to explore how response times to abortion and other survey questions may reveal other explanations for underreporting. And lastly, we explore the role of interviewer characteristics and past survey experience in women's willingness to report induced abortion. This work has three specific aims and associated hypotheses.

Aim 1: Estimate the prevalence of induced abortion overall and for subgroups using list experiment and direct questions. These estimates enable comparison of each methodology's performance overall and by subgroup. This aim is primarily descriptive but we also explore list experiment assumptions.

a. **Hypothesis 1a**: list experiment estimates will be significantly higher than those generated via direct questions on abortion.
b. **Hypothesis 1b**: the list experiment will reveal a significant design effect for some subgroup estimates. In other words, the addition of the sensitive item to the control list will significantly affect some individuals' responses to the control items, resulting in different propensities for respondents to

answer control items affirmatively across treatment and control groups. This is important because the validity of the list experiment estimates presumes no such design effect.

Aim 2: Assess failures in abortion reporting using response time paradata from direct and list experiment questions. This aim leverages the paradata available via the smart phone time stamped log, allowing us to examine the underlying mental processes involved in responding to cognitively demanding and sensitive survey questions.

a. **Hypothesis 2a**: response times on the list experiment directions comprehension question and the example list experiment question will be significantly shorter for respondents with more education.
b. **Hypothesis 2b**: response times on the treatment list experiment questions will be significantly longer for those who reported an abortion via the direct abortion questions.

Aim 3: Determine whether interviewer-respondent familiarity and respondent's prior survey experience are associated with improved reporting of abortion via direct abortion questions. The study uses local interviewers to implement the survey, thus interviewers are acquainted with many respondents. Additionally, the cross-sectional survey is administered repeatedly in communities every 6 to 12 months, thus some respondents are randomly selected for participation more than once. We aim to explore the potential impact of these survey design features on abortion reporting.

a. **Hypothesis 3a**: respondents who are acquainted with the interviewer will be more likely to report an abortion via the direct abortion questions.

b. **Hypothesis 3b**: respondents who participated in a previous survey round will be more likely to report an abortion via the direct abortion questions.

If the list experiment produces more valid estimates of abortion prevalence and incidence, its additional advantage can be rather easily incorporated into existing female population surveys in other settings. Investigation of the list experiment assumptions and associated paradata can provide insight into the performance of the list experiment questions. This knowledge can be applied to subsequent applications of the list experiment. Additional information on whether interviewer characteristics or respondent survey experience is associated with increased reporting of abortion can provide further information about the circumstances in which women may be more likely to report a past abortion. In combination, this dissertation seeks to explore abortion reporting and underreporting and determine whether better survey design can yield better data.

3. Research Design and Data

3.1 Data Source

Performance Monitoring and Accountability 2020 (PMA2020) is an interdisciplinary team from the Johns Hopkins Bloomberg School of Public Health (JHSPH) and several international universities and research organizations. In-country partners implement the data collection and spearhead dissemination efforts while the Bill & Melinda Gates Institute for Population and Reproductive Health at JHSPH provide overall direction and technical support. PMA2020 uses an innovative mobile-assisted technology to routinely collect data and update key family planning indicators every 6 to 12 months in 11 priority countries (Zimmerman et al. 2017). The data collection activities include a household, female, and service delivery point (SDP) survey. The mobile design enables rapid data collection and improved quality assurance, producing estimates that are comparable to those of Demographic and Health Surveys (Performance Monitoring and Accountability 2020 (PMA2020) ; Zimmerman et al. 2015). The platform is intended to measure progress towards achieving the Family Planning 2020 (FP2020) goal of providing contraception to 120 million additional women by the year 2020. Specifically, the project uses Open Data Kit (ODK) software on mobile phones to program the family planning surveys.

In each program country, a cadre of sentinel resident enumerators (REs) work in 100 to 250 nationally (or state or regionally) representative clusters to collect data at both the household and facility level. This sentinel area design allows PMA2020 to generate representative data on family planning program performance over time. REs transfer data by phone to a central server via the General Packet Radio Service (GPRS), a standard for wireless communication that is an efficient use of limited bandwidth (Lifewire 2017).

In addition to the recorded answers, the ODK platform collects data on the active screen time and short breaks spent on each ODK screen; this information is recorded in milliseconds. These data also capture when interviewers revisit questions, change answers, and violate survey constraints. These paradata provide information on the survey process whereas the survey data provide information on the family planning indicators of interest. In real-time, data are validated, aggregated, and prepared in tables, graphs, and maps.

Since its inception in 2013, PMA2020 has launched surveys with local partners in 11 countries, conducted 44 rounds of data collection, trained more than 1,800 REs, conducted more than 400,000 surveys, and produced and disseminated 40 family planning briefs to inform policy and program change in real time. We added the list experiment questions to one of the waves of the family planning survey conducted by PMA2020 in Rajasthan, India in 2017.

3.2 Study Population

Rajasthan has an estimated 2015 total population of more than 73 million. Recent estimates indicate an overall total fertility rate of 2.9 and a modern contraceptive prevalence rate of 62%, 76% of which is female sterilization (Office of the Registrar General & Census Commissioner et al. 2013). Induced abortion is broadly legal in India under the Medical Termination of Pregnancy Act of 1971 (Duggal and Ramachandran 2004). It is a commonly used means of fertility control, but its measurement has proven challenging (Arnold et al. 2002).

Official Rajasthan government statistics estimate that 22,980 induced abortions were conducted in 2013, resulting in an annual induced abortion incidence of approximately 2 per 1,000 women age 15 to 49 (Ministry of Health and Family Welfare 2013). These data do not capture the vast amount of inductions that occur in non-certified facilities in the private sector, or the substantial amount of self-induction. Results from a small study in Rajasthan found that 44% of women who reported a recent abortion sought induction services from a private sector doctor and 11% sought services from an informal or untrained practitioner, neither of which would be captured in official statistics (Jejeebhoy 2011). In addition, a small study in 2001 found that nearly 20% of women who reported a recent induced abortion initially attempted self-induction (Elul et al. 2004). With increasing availability of medical abortion drugs like misoprostol, which can be purchased from chemists without a prescription, the safety and success of self-induction in places like Rajasthan is increasing and thus a smaller proportion of inductions are being captured by facility data on induced abortion and post-abortion treatment of complications (Guttmacher Institute 2010). Singh *et al.* recently estimated a national abortion incidence of 47 abortions per 1,000 women of reproductive age, 73% of which were medical abortions that occurred outside the formal healthcare system (Singh et al. 2017).

A natural inclination is to seek the information from women themselves, but abortion remains highly stigmatized (Duggal and Ramachandran 2004). This is particularly true given the illegality of sex-selective abortion (Arnold et al. 2002). A recent community-based survey of women of reproductive age attempted to capture abortion data through direct questioning, but these estimates indicated that only 3% of recent pregnancies ended in abortion, (Office of the Registrar General & Census Commissioner et al. 2013). Another survey estimated that only 1% of women have ever induced an abortion (International Institute for Population Sciences (IIPS) and Macro International 2007, ICF International [Distributor], 2007). In light of these direct question estimates and the recent Singh *et al.* indirect estimates, it is widely believed that direct questioning results in substantial underestimation of induced abortion in Rajasthan (Singh et al. 2017). Obtaining valid data regarding induced abortion remains difficult in this setting, despite its legality.

3.3 Pilot Study

Prior to survey implementation, in-country partners aided in translating and back-translating the initial version of the list experiment questions to ensure correct wording in the local language (Hindi). Rajasthani colleagues and JHSPH personnel then conducted a small three-day pilot study. Seven REs from the Jaipur area who participated in the Round 1 Rajasthan data collection were invited to partake in the pilot survey prior to the broader Supervisor and RE trainings. On day one we began by conducting an informal focus group with the REs to solicit their input on how women talk about and refer to abortion, seeking to understand distinctions between induced abortion, miscarriage, and stillbirth. In particular, we discussed the wording for "had an abortion" and other euphemisms or phrases women might use to refer to induced abortion. We sought to ensure that respondents would understand the list experiment abortion item as induced abortion, not miscarriage or any other reproductive experience. By then end of the focus group, participants reached a consensus on how to refer to abortion in Hindi. This language distinguished between the intentionality of induced abortion as opposed to the passive nature of miscarriage. Following the focus group, we reviewed the list experiment and direct questions with the REs, paying particular attention to the translation. In addition, REs and study staff discussed and modified the list experiment control items.

On day two of the pilot, study staff and REs visited a peri-urban village on the outskirts of Jaipur and conducted 37 pilot surveys with available women of reproductive age. The pilot survey included the indirect and direct abortion questions in addition to some background and family planning questions. The end of the pilot survey included some quantitative cognitive interview questions to gauge understanding and difficulty related to the list experiment and direct abortion questions. After conducting these surveys, study staff and REs debriefed on the field experience and made updates to the list items, modified the translation, and expanded on the list experiment instructions to improve respondents' ability to understand and cognitively process the list experiment questions. The following day REs conducted another set of pilot surveys (47) in a different Jaipur village. This was followed by another field debrief and minimal additional abortion question modifications.

In the final iteration of the list experiment questions, 77% of pilot respondents interpreted and described the “had an abortion” item correctly on the subsequent cognitive interview question; this proportion was much lower than anticipated. However, only 9% described “had an abortion” incorrectly, while 14% found the cognitive interview question strange or difficult and did not know how to answer the question. Accounting for this non-response, 90% of respondents who actually answered the pilot cognitive interview question provided an accurate description of the “had an abortion” item. Despite the less than perfect respondent interpretation of “had an abortion”, the REs and in-country central staff determined this was the clearest Hindi phrasing and suggested that REs could use additional probing or hints to clarify any misunderstanding that arises with respondents.

We subsequently incorporated the final pilot version of the list experiment and direct abortion questions into the female questionnaire and associated training materials for Round 2. REs field-tested the female questionnaire containing the final list experiment and direct abortion questions during the refresher trainings.

3.4 Data Collection

The PMA2020 Rajasthan Round 2 survey aimed to sample approximately 6,000 women of reproductive age (15 to 49) residing in Rajasthan, India. Supervisor and RE trainings occurred in January and February 2017 and consisted of a four- to five-day refresher training. These refresher trainings covered new material, including the abortion questions, and a review of key aspects of the PMA2020 survey and protocol (survey staff had previously received 10 days of training in Round 1). Data collection occurred from March through May, 2017.

3.5 Sampling Design

PMA2020 utilized a two-stage cluster sampling design with urban/rural strata, and regions as the sampling domains. Individuals from the Indian Census Bureau then randomly select clusters from within the sampling domains using probability proportional to size, similar to the procedure utilized by the Demographic and Health Survey. In Rajasthan Round 1, which took place in summer 2016, 147 enumeration areas (EAs) were sampled from a complete sampling frame of EAs in the state using this process. These same 147 EAs were surveyed in Round 2. These EAs comprised 6 urban and 18 rural sampling domains. In each EA, REs listed and mapped all households. Central staff then randomly sampled 35 households from each EA listing. For Round 2, central staff again randomly sampled 35 households from each EA Round 1 sampling frame. REs also listed and mapped the private SDPs within each EA and supervisors randomly selected 3 in the event that an EA had more than 3.

Working with the local government, study staff also identified and surveyed the public SDPs that are assigned to serve selected EAs.

PMA2020 REs enumerated all members of the selected households through a household roster. REs then invited all women age 15 to 49 in the household to participate in a brief interview related to sexual and reproductive health and past pregnancies. Interested women were asked to consent in order to participate in the female interview. Women who were unable to respond for themselves (i.e. those with cognitive difficulty or who were ill) were excluded. The Institutional Review Boards (IRBs) at the Johns Hopkins Bloomberg School of Public Health (IBR No. 7571) and the India Institute of Health Management Research (IIHMR) provided ethical approval.

The list experiment lists used in the Round 2 female questionnaire are depicted in Table 3.1. The ODK form randomized half of the respondents to receive treatment list A (i.e. including the sensitive item) along with control list B (i.e. not including the sensitive item). The other half of the respondents received control list A and treatment list B. We added the list experiment questions to the first section of the survey following the background questions in order to limit women's ability to determine the intent behind these questions. All respondents received the control list followed by the treatment list in whichever group they were randomly assigned; previous work did not detect an order effect based on whether the treatment version of the list was seen first or second (Cowan et al. 2016).

We embedded the direct abortion questions in the reproductive history section following questions about previous births; this is consistent with the question order in India's National Family Health Survey (NFHS), enabling comparison of direct abortion question prevalence and incidence estimates. By putting the direct questions after the list experiment questions, we intended to eliminate the potential impact that answering a direct question on abortion might have on one's list experiment response. One study found that having direct questions prior to list experiment questions resulted in violations of some list experiment assumptions (including the assumption of no design effect) (Aronow et al. 2015). Other assumptions could be violated with list experiment questions first (Aronow et al. 2015), but we ultimately decided that the placement of the list questions at the outset of the survey followed by the direct questions in a later section would minimize unwanted context effects on list responses. This was further corroborated by feedback from other researchers' recent implementation of the list experiment to measure abortion (Moseson et al. 2015; Treleaven et al. 2017).

Table 3.1. Double list experiment items

	List A	**List B**	**Prevalence**
Item 1	Had a menstrual period	Used a sanitary pad during a menstrual period	High
Item 2	Used contraceptive injections	Used a female condom	Low
Item 3*	Had an abortion	Had an abortion	
Item 4	Visited a health facility or camp	Been visited by an anganwadi, ASHA, or other community health worker	High
Item 5	Had a C-section	Taken an ambulance to a hospital	Low

*Sensitive item; only added in treatment version of the list

3.6 Data

3.6.1 Aim 1

For Aim 1, we utilized the female questionnaire response data to calculate abortion prevalences using the list experiment and direct abortion questions. We also compared the list experiment and direct question estimates of lifetime experience of induced abortion. REs generated these data when they entered respondents' answers to specific questions into the smart phones.

3.6.2 Aim 2

The data used for Aim 2 leverages a novel data type that the smart phone captures in the process of conducting each PMA2020 survey. Technology on the phone generates a log of all interactions with the smart phone while implementing the survey. The information encoded in this log includes a record of the cumulative active screen time on each screen (including questions, warning messages, and instructions), the number of times the RE returns to a given screen, the number of times an RE changes a response, and the number of times an RE encounters a constraint as a result of illogical response combinations (e.g. currently age 20 and married at age 22). These survey process data, commonly referred to as paradata, enable in-depth exploration of the survey implementation experience. As such, it allows researchers to detect potentially cognitively challenging questions and other data quality issues, aspects of survey research that were previously largely unknown. In addition to these data generated in the course of implementing the female questionnaire, we utilized data on the RE characteristics that study staff collected following data collection. Using these three related data sources, we applied social and cognitive psychology theories with regard to social desirability pressure in order to explore what response times might reveal about potential cognitive difficulties and editing of responses.

3.6.3 Aim 3

For Aim 3, we utilized the female questionnaire response data and RE characteristics data to examine the role that respondent-RE acquaintance or prior PMA2020 survey experience may have on respondents' willingness to report induced abortion. In addition, we used the SDP data and the GPS data associated with the SDP and female data to quantify the abortion and post-abortion care service delivery environment for individual respondents.

We conducted all analyses in Stata version 15 and the R statistical platform (R Development Core Team 2015; StataCorp 2017).

4. Aim 1: Can a List Experiment Improve Validity of Abortion Measurement?

4.1 Background

Elective pregnancy termination, or induced abortion, is a frequent reproductive health experience that women throughout the world rely on to control their fertility. Current estimates indicate that approximately 56.3 (90% uncertainty interval [UI] 52.4 to 70.0) million induced abortions occurred annually from 2010 through 2014 (Sedgh et al. 2016). This corresponds to a global annual incidence of induced abortion of 35 (90% UI 33 to 44) per 1,000 women (Sedgh et al. 2016). Although induced abortion is common, measurement issues have long plagued this area of research, particularly in low-resource settings. Even in countries where abortion is broadly legal, facility data are often incomplete. In places where abortion is illegal or highly restricted, women often induce outside the formal health sector and only present at facilities in instances of perceived complications to receive post-abortion care (PAC). The availability of safer self-induction medicines, like misoprostol and mifepristone, have only increased the measurement challenge associated with induced abortion as evidence suggests that fewer women present at facilities for PAC (Guttmacher Institute 2010; Singh et al. 2017). Failing to accurately capture the extent to which women are relying on use of abortifacients purchased from pharmacies masks the need for reform with regard to medical abortion protocols and resources to improve availability of safe abortion services.

Relying instead on women's self-report of their experience with abortion presents a different set of issues, mainly that it is subject to substantial underreporting as a result of social desirability pressure. A face-to-face interview is among the most common means of survey administration, but even in legal contexts the pressure of social desirability results in less than 50% of abortions being reported (Jones and Kost 2007). Existing literature demonstrates that respondents are more willing to report sensitive behavior, including abortion, on self-administered questionnaires (Lensvelt-Mulders et al. 2005; Tourangeau and Yan 2007). However, in low-resource and low-literacy populations, trained enumerators typically administer surveys in a face-to-face setting. Researchers can employ audio computer-assisted self-interview (ACASI) in low-literacy populations, which often – although not always – results in higher reporting of sensitive sexual behaviors than face-to-face interviews (Langhaug et al. 2011; Lara et al. 2004).

Another way to reduce the pressure to respond in a socially desirable manner, regardless of the survey mode, is to ask about the sensitive item indirectly. Specifically, related to induced abortion, researchers have employed techniques such as randomized response technique (RRT) and the anonymous third-party reporting (ATPR) method, as well as a modification of ATPR referred to as the best friend or confidante method, with generally positive results (Coutts and Jann 2011; Elul 2004; Grossman et al. 2015a; Lara et al. 2004; Rossier et al. 2006; Yeatman and Trinitapoli 2011). Additionally, researchers in India have used a mixed methods narrative approach to improve reporting with significant success (Edmeades et al. 2010). Each of these methods has a number of strengths and limitations but each typically results in more valid estimates of induced abortion than direct questions, although there are many exceptions (Elul 2004; Fuentes 2017; Moseson et al. 2017b). Ultimately there is no gold standard for abortion measurement and the methodological choice is driven by the context, the research objective(s), and the budget (Rossier 2003).

In India, induced abortion has been legal on request for a broad set of circumstances since the passage of the Medical Termination of Pregnancy Act of 1971. In Rajasthan, India specifically, official statistics from the Ministry of Health and Family Welfare indicate that 22,980 induced abortions were conducted in 2013, resulting in an annual induced abortion incidence of approximately 2 per 1,000 women age 15 to 49 (Ministry of Health and Family Welfare 2013). However, these data are incomplete as they only include terminations conducted in certified facilities that are registered with the government to provide abortion. This excludes many private sector facilities as the process of registering with the government is cumbersome, while other providers are unaware of this system. Results from a small study of women in Rajasthan revealed that 44% of women who reported a recent abortion had gone to a private sector doctor and 11% utilized informal or untrained providers; these abortions would largely be unaccounted for in government statistics (Jejeebhoy 2011).

Government abortion service data also omit the substantial occurrence of self-induction. A small survey of Rajasthani women conducted in 2001 found that nearly

1 in 5 women who reported a recent induced abortion had initially attempted self-induction (Elul et al. 2004). The availability of misoprostol and mifepristone at pharmacies and chemists has only grown in recent years. Private sector drug distribution data indicate that the availability of misoprostol-only drug sales to wholesalers in India increased 646% from 2002 to 2007 (Fernandez et al. 2009). More recent estimates indicate that the volume of medical abortion drug sales is equivalent to approximately 34 abortions per 1,000 women age 15 to 49 nationwide (Singh et al. 2017).

The current chapter applies a novel method, known as the list experiment, to try to improve measurement of induced abortion in Rajasthan, India. Social psychologists originally developed the list experiment methods in the 1980s in an attempt to elicit more truthful responses to sensitive questions (Kuklinski et al. 1997; Miller 1984; Sniderman and Carmines 1997). We describe the design of the list experiment in detail in Chapter 2. We employed the double list experiment in order to improve the efficiency of the estimator, whereby every respondent received a treatment version of one list and a control version of another list. Thus, everyone served as control and treatment within the sample.

For the list experiment to yield unbiased estimates of a given sensitive survey item, three assumptions must be met: 1) effective randomization, i.e. treatment and control groups are the same; 2) no design effect, i.e. addition of sensitive item to the treatment list does not affect responses to the control items; and 3) honest responses (Blair and

Imai 2012; Glynn 2013). Assumption 1 is under investigator control, whereas Assumptions 2 and 3 can be violated if respondents evaluate items on the list relative to one another or respond in a non-accurate manner, either intentionally or not (e.g. due to misunderstanding). Use of the list experiment resulted in generally improved measurement of sensitive attitudes or behaviors compared to direct question results in a number of contexts (Aronow et al. 2015; Comșa and Postelnicu 2013; Gonzalez-Ocantos et al. 2012; Tsuchiya et al. 2007; Wolter and Laier 2014), but has also performed poorly in other applications, producing estimates not significantly more valid than those generated via other indirect methods or direct questioning (Biemer and Brown 2005; Coutts and Jann 2011; Droitcour et al. 2004; Rosenfeld et al. 2014). As a result, social scientists have not reached a consensus on the effectiveness of this method in measuring sensitive items. In recent years, researchers have begun to assess its performance specifically with regard to abortion measurement (Cowan et al. 2016; Moseson et al. 2015; Moseson et al. 2017a; Moseson et al. 2017b; Treleaven et al. 2017).

The list experiment has recently been used to measure induced abortion with relative success in various contexts. In a 2015 study measuring lifetime experience of abortion in Liberia, results indicated that 32% of women had ever had an abortion (Moseson et al. 2015). The list experiment estimate was five times greater than the only previous comparable estimate of induced abortion in the country, which had been estimated via direct survey questions. More recent research in the United States piloted list experiment questions using an online convenience sample of 1,233 women (Cowan et al. 2016). Twenty-two percent and 18% of women reported a prior induced abortion in response to list experiment and direct questions, respectively; however, these estimates were not statistically significantly different (Cowan et al. 2016). Other unpublished work has also been conducted in Vietnam and Texas with mixed results (Moseson et al. 2017b; Treleaven et al. 2017).

The current study aims to improve upon this previous research and more rigorously evaluate the use of a list experiment in the measurement of induced abortion in a low-resource setting. We added list experiment questions to a population-based survey of reproductive age women in Rajasthan, India and compared resulting estimates to those we obtained via direct questioning in the same sample. We then evaluated list experiment assumptions, providing additional evidence as to the quality of the data collected from this technique in low-resource settings.

4.2 Methodology

4.2.1 Data

To investigate performance of the list experiment in measuring induced abortion, researchers from the Bill & Melinda Gates Institute for Reproductive Health at the Johns Hopkins Bloomberg School of Public Health (JHSPH) and the Indian Institute of Health Management and Research (IIHMR) added list experiment questions to a family planning survey of reproductive age women in Rajasthan, India. This survey was part of the PMA2020 project, which uses an innovative mobile-assisted technology to routinely collect data and update key family planning indicators every

6 to 12 months in 11 priority countries (Zimmerman et al. 2017). The platform is intended to measure progress towards achieving the Family Planning 2020 (FP2020) goal of providing contraception to 120 million additional women by the year 2020. Additional details are provided in Chapter 3.

Investigators first conducted a three-day pilot with seven resident enumerators (REs) in order to adapt the list experiment and direct abortion questions and associated translations. During the pilot, we determined the correct Hindi translation for induced abortion. All REs then received a five-day refresher training that reiterated important elements of survey implementation and the core family planning survey, and presented the new abortion material. The REs had received ten days of training prior to conducting the initial family planning survey in 2016. Data collection for the second round, which contained the abortion questions, occurred in April and May of 2017. PMA2020 uses a probabilistic two-stage cluster sampling design using urban/rural and region as the sampling domains and probability proportional to size sampling for the selection of enumeration areas (EAs) within sampling domains. Investigators took a random sample of 35 households from each of the 147 EA sampling frames created in Round 1. All eligible women, i.e. those age 15 to 49, in selected households were invited to participate in a brief interview related to sexual and reproductive health and past pregnancies. REs requested consent from all participants prior to administering the survey. The Institutional Review Boards (IRBs) at JHSPH and IIHMR provided ethical approval.

We used the ODK form to randomize half of the respondents to receive treatment list A (i.e. including the sensitive item) along with control list B (i.e. not including the sensitive item). The other half of the respondents received control list A and treatment list B. We placed the list experiment questions in the first section of the survey following background questions in order to limit women's ability to determine the intent behind these questions. We then embedded the direct abortion questions in the reproductive history section following questions about previous births; this is consistent with the question order in India's National Family Health Survey (NFHS). Putting the direct questions after the list experiment questions was also intended to eliminate the potential impact that answering a direct question on abortion might have one's list experiment response.

4.2.2 Analysis

Given the indirect nature of the list experiment, the associated analyses are not straightforward. Suppose there is a population of n respondents partitioned into n_1 respondents who encountered the control list and n_2 respondents who received the treatment list. Let y_{Ai+} equal the total number of items that individual *i* reported ever experiencing from the treatment version of list A, which includes the sensitive item, and y_{Ai-} equal the number of items that individual *i* reported ever experiencing from the control version of list A, which does not include the sensitive item. The estimated proportion ever experiencing an induced abortion using list A, $\hat{\pi}_A$, is given by equation 1. A similar expression can be used for list B as seen in equation 2. These expressions allow one to calculate a difference in means between the average item count responses on the treatment and control versions of the lists.

$$\hat{\pi}_A = \frac{1}{n_2}\sum_{i=n_1+1}^{n} y_{Ai+} - \frac{1}{n_1}\sum_{i=1}^{n_1} y_{Ai-} \quad (1)$$

$$\hat{\pi}_B = \frac{1}{n_1}\sum_{i=1}^{n_1} y_{Bi+} - \frac{1}{n_2}\sum_{i=n_1+1}^{n} y_{Bi-} \quad (2)$$

Because we exposed each respondent to two lists in the double list experiment, we used the sample to produce two estimates of induced abortion prevalence using equations 1 and 2 above. We then took the average of these two induced abortion prevalence estimates for list A and B.

The assumption of effective randomization can be quantitatively assessed by conducting chi-squared tests to determine if the distribution of respondents by a socioeconomic characteristic is statistically significantly different by treatment assignment.

The assumption of no design effect, in other words that the addition of the sensitive item does not change an individual's response to the control items, can be represented mathematically. Using potential outcomes notation, we let $Z_{ij}(t)$ denote a binary variable that represents respondent *i*'s response for each control item *j* for *j* = 1,…, *J*, where *J* is the total number of control items on the list, under treatment status *t* = 0 for control list and 1 for treatment. Thus, for each *i* = 1,…, *N*, if there is no design effect, $\sum_{j=1}^{J} Z_{ij}(0) = \sum_{j=1}^{J} Z_{ij}(1)$.

The assumption that respondents are not falsifying their responses can also be represented mathematically, where the observed response of individual *i* for item *J*+1, the sensitive item, is assumed to be equal to the truthful answer to the sensitive question, i.e. $Z_{i,J+1}(1) = Z^{*}_{i,J+1}$, where $Z^{*}_{i,J+1}$ represents a truthful response to the sensitive item; $Z_{i,J+1}(0)$ is not defined since the sensitive item is not included in the control list. The implications of these assumptions are essentially equivalent in application.

The potential answer respondent *i* would give under the treatment and control conditions is denoted by $Y_i(0) = \sum_{j=1}^{J} Z_{ij}(0)$ and $Y_i(1) = \sum_{j=1}^{J+1} Z_{ij}(1)$, respectively, where $Y_i(1)$ is an integer between 0 and *J*+1 and $Y_i(0)$ is an integer between 0 and *J*. The observed response for respondent *i* is denoted $Y_i = Y_i(T_i)$ where T_i denotes the treatment status actually assigned, 0 or 1.

The design effect and no falsified answers assumptions can be assessed by investigating the conditional probabilities of reporting *y* number of items depending on treatment assignment *T*, where the null hypothesis can be expressed as:

$\Pr(Y_i \leq y | T_i = 0) \geq \Pr(Y_i \leq y | T_i = 1)$ for all *y* = 0,…, *J* – 1, and (3)

$\Pr(Y_i \leq y | T_i = 1) \geq \Pr(Y_i \leq y - 1 | T_1 = 0)$ for all *y* = 1,…, *J*.

If baseline responses are honest and respondents never over-report the sensitive behavior, then one can test whether the joint proportion, θ_y, defined as $\theta_y = \Pr(Y_i \geq y|T_i = 0) - \Pr(Y_i \geq y|T_i = 1)$, is significantly different from zero in the negative direction. If the assumption of no design effect is satisfied, the addition of the sensitive item to the control list will make the response variable Y_i in the treatment group larger than the control group response variable (the first line of equation 3) but by no more than one item (the second line of equation 3). If one of these joint proportions is negative, the assumption of no design effect (i.e. that the addition of the sensitive item to the control list does not affect an individual's response to the control items) is necessarily violated, as is the assumption of honest responses. Thus to assess whether the assumptions of no design effect and no falsified responses was violated, we conducted one-sided t-tests for the sample overall and by subgroup to determine if any of the θ_y were significantly less than 0 (Glynn 2013). If θ_ys are less than 0, one can re-estimate the proportion experiencing the sensitive item by truncating the θ_y at 0 as seen in equation 4:

$$\theta_y = \theta_y \text{ if } \theta_y \geq 0, \text{ and} \tag{4}$$

$$\theta_y = 0 \text{ if } \theta_y < 0$$

Then one can sum across the estimated θ_ys in the proportion reporting at least y items, which provides the piecewise list experiment estimate.

In this chapter, we first calculated the socioeconomic characteristics of the sample overall and by treatment assignment. We also included the p-value associated with the chi-squared test of difference to determine if the assumption of effective randomization holds. We then generated list experiment and direct lifetime experience of abortion estimates overall and by subgroup. Subgroups included age groups, marital status, educational attainment, wealth quintile, caste, religion, residence (urban/rural), and parity. When calculating the overall and subgroup prevalences, we investigated and adjusted for violations of these aforementioned assumptions using the piecewise estimator previously described (Blair and Imai 2012; Glynn 2013).

To calculate the standard errors (SEs) associated with the list experiment estimates and to generate 95% confidence intervals (CIs) around the estimated difference between the list experiment and direct estimates in making statistical inferences, we used a resampling method. Specifically, we used the independent and identically distributed (iid) bootstrap with bias corrected CIs to account for potential non-normality of the bootstrapped distribution of estimates (Carpenter and Bithell 2000; Efron 1987).

Given the complex sampling design, we employed a two-stage resampling procedure to generate the sample distribution, accounting for the strata (urban/rural) and then selecting a random sample of n clusters (EAs) with replacement from the n sample clusters. The random sample of m_i elements within the i^{th} sample cluster was maintained, including all women within a given cluster each time it was randomly selected (StataCorp 2015a). Given the number of units within clusters varied, the overall sample size across the resamplings also varied. For each of the samples, the survey weights, which accounted for the design weight and non-response, were normalized so that the average of the weights was equal to 1.0.

We resampled 1,000 times for each estimate, generating the sampling distribution of piecewise estimates for the list experiment as well as the difference between the piecewise estimates and the direct estimates, overall and by subgroup (Efron and Tibshirani 1994). The bias corrected 95% CI was used to account for the potential non-normality of the abortion estimate. Abortion is rare and bound by 0, and has a potentially non-stable standard error. This bootstrap method generates one constant that transforms the sampling distribution in a manner that is normalizing and variance stabilizing (Efron 1987). With this method, it is assumed that normality and constant standard error can be achieved by some transformation g, such that $\hat{\phi} = g(\hat{\theta})$, $\phi = g(\theta)$, and $\frac{\hat{\phi}-\phi}{\tau} \sim N(-z_0, 1)$, where z_0 is the bias constant and τ is the constant standard error of $\hat{\phi}$. We used this method to test whether the difference between the overall and subgroup abortion estimates from the direct questions and the list experiment questions were statistically significantly different.

We conducted all analyses in Stata version 15 and the R statistical platform, incorporating survey weights and accounting for the complex sampling design (R Development Core Team 2015; StataCorp 2015b).

4.3 Results

The final sample of women who completed the female questionnaire included 6,035 women. The response rate was 97.8%. Table 4.1 presents the characteristics of Rajasthani women age 15 to 49. On average, women were 29 years old, the majority of whom (75.7%) were currently married or cohabiting. Most women had never attended school (36.8%), were of an other backward caste (39.2%), were Hindu (85.3%), and resided in rural areas (64.2%). Nearly one-third (31.1%) of women were nulliparous, while 36.1% had 1 to 2 children and 24.7% had 3 to 4 children; only 8.2% had 5 or more children.

Table 4.2 contains the overall list experiment estimate of lifetime experience of abortion among all women using the piecewise estimator, by list. The simple difference in means estimator generated an abortion prevalence of 0.2% and -1.4% on list 1 and 2, respectively. Accounting for clear violations of the list experiment assumptions, namely a potential design effect whereby the joint proportion (Row 5) is negative, generated estimates of 2.5% and 1.1% on list 1 and 2, respectively. As such, the overall piecewise estimate of lifetime prevalence of abortion was 1.8% (Table 4.2).

Using this piecewise estimator approach, we present the associated list experiment abortion prevalence estimates overall and by background characteristics, along with the direct abortion estimates, in Table 4.3. The direct and list experiment estimates displayed similar trends within some background characteristics. For instance, abortion prevalences increased with age, peaking at 30 to 39 before reducing among the oldest cohort of women; prevalence estimates generated via direct and list experiment questions increased with increasing education but declined slightly among those with higher or postgraduate education; Hindu women reported the lowest experience with abortion while Muslim women reported slightly higher levels of abortion and other religions' estimates were higher still; and abortion estimates were higher among urban compared to rural women. However, direct and list experiment estimates followed somewhat divergent patterns within other background characteristics. For example, the direct estimates of abortion were higher for currently married or cohabiting women while list experiment estimates were slightly higher for divorced, separated, or widowed women; direct abortion estimates increased with increasing wealth while list experiment estimates remained similar across wealth quintiles; and estimates by caste and parity followed different patterns.

When comparing the direct and list experiment estimates quantitatively, very few estimates were statistically significantly different (Table 4.3). Overall, the list experiment abortion prevalence estimate was 1.7% lower than the direct estimate, which was statistically significant ($p<0.01$). The only subgroup estimates that were significantly different by method were currently married or cohabiting women (2.4%, $p<0.01$), the second wealthiest (2.2%, $p<0.05$) and wealthiest (3.3%, $p<0.05$), and urban women (2.9%, $p<0.05$), all of which had significantly higher direct estimates.

Examining the list experiment assumptions begins to reveal potential explanations for its poor performance. Table 4.4 presents the background characteristics of the women by treatment list, providing clear evidence that we achieved effective randomization, as anticipated. The only statistically significant difference was the average age by treatment list (29.3 versus 28.6, p=0.017), but this was not likely to result in qualitative differences between the groups' responses. However, results from the assessment of design effects demonstrated clear violations of the list experiment assumption, and likely the assumption of honest responding. Table 4.5 contains the p-values for the associated design effect significance test overall and by background characteristics for each list. By this metric, list 1 appeared to have performed better, with evidence of significant design effects only for those who never attended school and those from an other backward caste (Table 4.5). List 2 had more issues with regard to design effects, with significant violations detected among women aged 15 to 19, those from a scheduled tribe, Hindu women, nulliparous women, and women who reported no past experience with abortion via the direct questions.

Investigating the list experiment performance further by whether women reported an abortion via the direct questions provides additional insights (Tables 4.6a and 4.6b). Among women who reported having an abortion on the direct questions, the list experiment estimate of abortion prevalence was 41.1% among those who received treatment list 1 and 94.5% among those who received treatment list 2; the average estimate was 67.8% (Table 4.6a). This provides clear evidence that women were actually *less* likely to include their experience with abortion in their numeric response to the list experiment questions than they were on the direct questions asked later in the survey; this was particularly true for list 1. Note that there were no negative joint proportions (Table 4.6a, Rows 5) in the piecewise estimates among women with a known abortion.

Among women who reported no abortion on the direct questions, women were again less likely to report this experience via the list experiment, resulting in a difference in mean estimate of -1.1% and -5.1% for list 1 and 2, respectively (Table 4.6b, Rows 5). Results from the piecewise estimator generated estimates of 1.4% and 0.4%, respectively, thus identifying a small proportion (0.9%) of women who had an abortion as per the list experiment estimates but who denied it on the direct abortion questions.

4.4 Discussion

Results from this chapter provided list experiment and direct question estimates of lifetime experience of induced abortion and evaluated the performance of the list experiment methodology. Overall, 1.8% of women reported a past abortion via the list experiment questions whereas 3.5% of women reported an abortion via the direct questions; this difference was statistically significant. As such, the list experiment failed to produce more valid estimates of this sensitive behavior on a population-based survey of reproductive age women in Rajasthan, India. Estimates within subgroups were generally similar across methodologies, but the direct abortion estimates were significantly higher among currently married or cohabiting women, wealthier women, Hindu women, and urban women.

These results were not entirely unexpected given that recent applications of the list experiment to measure aspects of induced abortion in low-resource settings have generated mixed results. The list experiment performed well in the initial Liberia application that measured lifetime experience of abortion (Moseson et al. 2015), while it produced lower than expected sex selective abortion estimates in Vietnam (Treleaven et al. 2017). Additionally, application of the list experiment produced only slightly higher estimates in a US online survey (Cowan et al. 2016) and much higher than anticipated estimates of self-induced abortion in Texas (Moseson et al. 2017b). We should also note that attempts to measure induced abortion *incidence* via a double list experiment on our survey of Rajasthani women failed to produce a positive incidence estimate using the difference in means calculation (-34.10 abortions per 1,000 women age 15 to 49), but the piecewise estimated abortion incidence was 15.82 abortions per 1,000 women of reproductive age (95% CI 4.98-28.13); this was significantly higher than the direct abortion incidence estimate of 4.14 per 1,000 (95% CI 1.79-6.49). Taken together, these findings cast doubt on the reliability of using a list experiment to measure sensitive behaviors in a low-resource setting.

In the context of the PMA2020 Rajasthan survey, one obvious explanation for the poor list experiment performance is the evidence of assumption violations. Women may have evaluated the list items relative to one another resulting in treatment list responses to control items being different than the corresponding responses on the control list. This presentation of a design effect was partially accounted for in the piecewise estimator, but the associated estimates would not fully adjust for this behavior. This behavior specifically among women who would have either responded with the lowest number of items on the treatment list (0) or the highest (5) would be an example of a floor or ceiling effect, which may present a violation of the assumption of honest responding. Alternatively, women may have simply not included their abortion in their treatment list response. This is readily apparent when examining the list experiment results among the subgroups of women who did and did not report an abortion on the direct questions.

There are several other potential explanations for why the list experiment failed to produce improved estimates of induced abortion. The list experiment may simply be too cognitively demanding for respondents. This may lead respondents to provide spurious answers to the list experiment questions. Alternatively, poor cognitive ability or numeracy may have resulted in women incorrectly providing responses regarding the specific items they had experienced, which is a clear violation of the confidentiality that this method is meant to afford. Although our direct and list experiment estimates of abortion by education were not statistically significantly different, there *was* evidence of a significant design effect among the subgroup of respondents who had never attended school. This provides some support to the poor cognitive ability hypothesis.

Beyond potential poor understanding of the list experiment questions and associated instructions, women may have not interpreted the sensitive item, or their corresponding past behavior, accurately. To the extent that a woman does not view a past experience as an abortion, she will correctly not include that experience in her answer to questions about abortion on surveys (Kanstrup et al. 2017; Simonds et al. 1998). Relatedly, our phrasing of induced abortion may have been too narrow for the Indian context. Evidence from other places highlight the experience of simply “bringing back one’s period”, similar to menstrual regulation (Plummer et al. 2008; Rahman et al. 2014). This may help to explain the low direct and list experiment estimates, but not the poor list experiment performance relative to direct questions.

Additionally, given the placement of the list experiment at the beginning of the survey and the direct questions later, more rapport between REs and respondents may have developed between when REs asked the list experiment questions and the direct questions (Sudman et al. 1996). As such, women may have felt more comfortable revealing their abortion later in the survey, or the initial exposure to the topic of abortion in the list experiment cognitively primed the respondent (Sudman et al. 1996). Lastly, the poor list experiment performance could simply be a result of poor implementation on the part of the REs or poor engagement on the part of the respondents. Next steps include further investigation of some of these proposed explanations for the list experiment’s failure to produce improved estimates of abortion prevalence or incidence in this sample of Rajasthani women.

This study has a number of strengths. The data collected provide a large, representative sample of reproductive age Rajasthani women. The female questionnaire included both direct and indirect (list experiment) questions on abortion, providing an in sample contemporaneous comparison of the two methodologies, and the large sample size provided sufficient power to detect significant differences across subgroups of background characteristics. Additionally, interviewers were largely *resident* enumerators, meaning that the interviewer had the potential added advantage of being from the area of most of her respondents. This may have improved survey implementation and translation into local languages, and evidence suggests it may have helped to create an environment in the survey interaction that increased respondents' willingness to reveal sensitive behaviors (Rodriguez et al. 2015; Sana et al. 2016; Weinreb 2006).

Despite these strengths, this investigation had a number of limitations. These limitations present opportunities for improvement in future list experiment implementations and we present them as a practical set of lessons learned. Our list experiment design used four control items, but we recommend trying fewer control items, which may minimize the cognitive demand of the question. The success of the list experiment in Liberia may be partially attributable to their use of three control items (Moseson et al. 2015). We also recommend doing extensive testing of different control items to identify the best performing control lists. We generated the control lists in conjunction with our in-country partners and made several modifications during the pilot, but a larger pilot that tested more control items may have led to improved list experiment performance. Additionally, we highly recommend conducting qualitative cognitive interviews during the pilot to better understand how the respondents are interpreting the list experiment and the individual items. We only added quantitative cognitive interview questions at the end of the pilot questionnaire, limiting our ability to determine respondents' understanding of the list experiment and whether they knew it protected the confidentiality of their responses.

We encourage the use of a dummy list following the list experiment instructions to ensure the respondent knows to provide only a numeric response, which we did using a list of local foods. Even better would be the use of a dummy list that measures something innocuous that is measured directly elsewhere in the survey, like whether the respondent has ever had a child. This will determine whether the list experiment can effectively be used to measure any item accurately in the given survey context, thus revealing whether a failure of the abortion list experiment questions is due to the sensitive nature of abortion or poor performance of the list more generally.

Regarding training, adequate list experiment training time must be scheduled regardless of interviewers' prior survey participation. The REs we trained were not professional interviewers and 38.5% of REs indicated on a post-data collection survey that they experienced difficulty implementing the list experiment questions as intended. Additional training may have mitigated this difficulty. Lastly, we had no external estimates against which to validate the direct or list experiment results. However, we recommend including direct questions for comparison as we did.

Further examination is required to determine contexts and conditions in which application of a list experiment is most likely to be beneficial and result in improved abortion estimates. A recent publication more thoroughly summarizes best practices and remaining questions regarding using a list experiment to measure induced abortion (Moseson et al. 2017a). We encourage those who recently used or are planning to use a list experiment to publish their findings regardless of the list experiment's performance. Subsequently conducting a meta-analysis of list experiment performance in measuring abortion, or other sensitive behaviors, will allow advancement of the science around this methodology and failures of the list experiment, like this one, need to be represented. Time will tell whether we can learn enough about the list experiment to effectively and consistently leverage its potential benefits to improve measurement of induced abortion, or whether the methodology will lose appeal among abortion researchers.

Table 4.1. Socioeconomic characteristics of Rajasthani women age 15 to 49[1]

	Total	
	%	N[2]
Mean age (SE)	28.9 (0.16)	6,035
Marital status		
Currently married/cohabiting	75.7	4,557
Divorced or separated/widowed	2.7	162
Never married	21.6	1,302
School		
Never attended	36.8	2,221
Primary	24.3	1,469
Secondary	17.6	1,059
Higher or postgraduate	21.3	1,285
Wealth		
Poorest	16.5	997
Second poorest	17.5	1,056
Middle	19.7	1,186
Second wealthiest	21.5	1,295
Wealthiest	24.9	1,500
Caste of household head		
Scheduled caste	22.3	1,346
Scheduled tribe	17.3	1,042
Other backward caste	39.2	2,362
General	21.2	1,279
Religion of household head		
Hindu	85.3	5,148
Muslim	13.3	801
Other	1.4	86
Residence		
Rural	64.2	3,874
Urban	35.8	2,160
Parity		
0	31.1	1,873
1-2	36.1	2,177
3-4	24.7	1,487
5+	8.2	493
Abortion (direct question)		
No	96.5	5,823
Yes	3.5	211
Total	100.0	6,035

[1]Estimates and Ns weighted
[2]Ns do not always add to 6,035 due to missing

Table 4.2. List experiment estimates of lifetime experience of abortion among all Rajasthani women age 15 to 49 using the piecewise estimator

		Number of reported items (proportion)						
List 1	Source	0	1	2	3	4	5	Sum
Row 1	List with abortion	0.052	0.474	0.399	0.065	0.009	0.002	1.000
Row 2	Proportion at least	1.000	0.948	0.474	0.075	0.011	0.002	--
Row 3	List without abortion	0.042	0.470	0.429	0.055	0.004	0.000	1.000
Row 4	Proportion at least	1.000	0.958	0.488	0.059	0.004	0.000	--
Row 5	Row 2 minus Row 4	0.000	-0.010	-0.014	0.017	0.007	0.002	0.002
Row 6	Exclude violations	0.000	0.000	0.000	0.017	0.007	0.002	0.025
List 2	Source							
Row 1	List with abortion	0.115	0.529	0.277	0.068	0.008	0.002	1.000
Row 2	Proportion at least[1]	1.000	0.885	0.356	0.078	0.010	0.002	--
Row 3	List without abortion	0.119	0.505	0.293	0.078	0.005	0.000	1.000
Row 4	Proportion at least	1.000	0.881	0.376	0.083	0.005	0.000	--
Row 5	Row 2 minus Row 4	0.000	0.004	-0.020	-0.004	0.005	0.002	-0.014
Row 6	Exclude violations	0.000	0.004	0.000	0.000	0.005	0.002	0.011
Average estimate across lists								-0.62%
Average estimate across lists, violations excluded								1.80%

Note: Rows 1 and 3 represent the proportions reporting each number of items on the treatment and control lists, respectively. Rows 2 and 4 represent the proportions reporting *at least* each number of items on the treatment and control lists, respectively. Row 5 represents the difference between Row 2 and 4, which is equal to the proportion of women who report having an abortion and the total number of treatment list items indicated by the column (i.e. the joint proportion). Row 6 is a replicate of Row 5 where negative estimates have been excluded. The sum column for Row 5 and 6 represent the overall estimate of the proportion of women reporting a past abortion, including and excluding violations (i.e. negative joint proportions in Row 5), respectively.

Table 4.3. Estimate of lifetime experience of induced abortion among Rajasthani women age 15 to 49, by socioeconomic characteristics and measurement methodology[1]

	Direct	List Experiment	List Experiment - Direct
	% (SE)	% (SE)	% (95% CI)[2]
Age			
15-19	0.2 (0.1)	0.7 (1.0)	0.5 (-0.2, 1.7)
20-29	3.9 (1.0)	4.3 (1.6)	0.4 (-2.4, 3.1)
30-39	5.7 (0.9)	5.3 (1.9)	-0.4 (-4.3, 2.6)
40-49	2.8 (1.0)	1.2 (1.3)	-1.6 (-3.2, 0.4)
Marital status			
Currently married/cohabiting	4.6 (0.9)	2.2 (1.0)	**-2.4 (-3.9, -1.1)**
Divorced or separated/widowed	1.4 (1.1)	3.4 (5.1)	1.9 (-1.3, 2.9)
Never married	0.1 (0.1)	0.7 (0.9)	0.6 (-0.3, 1.5)
School			
Never attended	2.8 (0.6)	2.2 (1.6)	-0.6 (-2.8, 1.8)
Primary	3.4 (0.8)	2.4 (1.1)	-1.0 (-3.5, 1.2)
Secondary	4.4 (1.4)	5.5 (2.3)	1.2 (-3.6, 5.5)
Higher or postgraduate	4.2 (1.0)	2.3 (1.8)	1.9 (-4.4, 1.1)
Wealth			
Poorest	1.5 (0.6)	2.2 (1.4)	0.7 (-1.3, 3.1)
Second poorest	1.3 (0.4)	2.6 (1.7)	1.3 (-0.9, 3.2)
Middle	3.6 (1.0)	2.7 (1.3)	-0.9 (-3.3, 0.8)
Second wealthiest	4.6 (1.5)	2.4 (1.6)	**-2.2 (-5.0, -0.4)**
Wealthiest	5.3 (1.3)	2.1 (1.7)	**-3.3 (-7.2, -1.3)**
Caste of household head			
Scheduled caste	3.6 (1.1)	2.4 (1.7)	-1.3 (-4.3, 0.7)
Scheduled tribe	3.6 (1.4)	6.6 (2.3)	2.9 (-1.6), 7.3)
Other backward caste	3.0 (0.9)	2.0 (1.3)	-1.0 (-3.4, 1.1)
General	4.2 (1.1)	4.1 (1.8)	-0.1 (-5.1, 2.7)
Religion of household head			
Hindu	3.3 (0.6)	1.9 (0.9)	-1.4 (-3.4, 0.2)
Muslim	4.4 (2.1)	3.4 (3.8)	-1.0 (-8.1, 1.8)
Other	9.0 (3.1)	12.4 (40.0)	3.4 (-10.6, 97.1)
Residence			
Rural	1.9 (0.4)	1.3 (0.7)	-0.6 (-1.9, 0.7)
Urban	6.4 (1.7)	3.6 (1.8)	**-2.9 (-7.0, -0.5)**
Parity			
0	0.2 (0.1)	1.2 (1.0)	0.9 (-0.1, 2.1)
1-2	5.5 (1.1)	3.2 (1.4)	-2.3 (-5.8, 0.4)
3-4	4.8 (1.0)	2.1 (1.7)	-1.8 (-4.7, 0.8)
5+	3.2 (1.1)	6.6 (3.1)	3.4 (-2.9, 8.0)
Total	3.5 (0.7)	0.0 (0.8)	**-1.7 (-3.3, -0.5)**

[1]Estimates weighted

[2]Bias-corrected bootstrapped 95% confidence intervals

Table 4.4. Socioeconomic characteristics of Rajasthani women age 15 to 49, by list[1]

	List 1		List 2		Total		p-value
	%	N	%	N	%	N	
Mean age (SE)	29.3 (0.24)	2,989	28.6 (0.18)	3,046	28.9 (0.16)	6,035	**0.017**
Marital status							
Currently married/cohabiting	76.6	2,283	74.8	2,274	75.7	4,557	0.339
Divorced or separated/widowed	2.5	75	2.9	87	2.7	162	
Never married	20.9	624	22.3	679	21.6	1,302	
School							
Never attended	37.6	1,122	36.1	1,098	36.8	2,221	0.575
Primary	23.5	703	25.2	766	24.3	1,469	
Secondary	17.5	522	17.6	537	17.6	1,059	
Higher or postgraduate	21.5	641	21.1	644	21.3	1,285	
Wealth							
Poorest	16.1	483	16.9	515	16.5	997	0.783
Second poorest	17.1	510	17.9	546	17.5	1,056	
Middle	19.8	591	19.5	595	19.7	1,186	
Second wealthiest	21.9	654	21.0	641	21.5	1,295	
Wealthiest	25.1	751	24.6	750	24.9	1,500	
Caste of household head							
Scheduled caste	22.2	663	22.4	683	22.3	1,346	0.600
Scheduled tribe	17.9	534	16.7	508	17.3	1,042	
Other backward caste	38.4	1,146	39.9	1,216	39.2	2,362	
General	21.5	642	20.9	637	21.2	1,279	
Religion of household head							
Hindu	84.5	2,526	86.1	2,623	85.3	5,148	0.248
Muslim	13.9	415	12.7	386	13.3	801	
Other	1.6	48	1.2	38	1.4	86	
Residence							
Rural	64.3	1,920	64.1	1,954	64.2	3,874	0.930
Urban	35.7	1,068	35.9	1,092	35.8	2,160	
Parity							
0	30.7	917	31.4	955	31.1	1,873	0.429
1-2	35.2	1,052	37.0	1,125	36.1	2,177	
3-4	25.5	762	23.8	725	24.7	1,487	
5+	8.5	254	7.9	239	8.2	493	
Abortion (direct question)							
No	96.7	2,889	96.3	2,935	96.5	5,823	0.499
Yes	3.3	100	3.7	111	3.5	211	
Total	100.0	2,989	100.0	3,046	100.0	6,035	

[1]Estimates and Ns weighted

Table 4.5. Detection of list experiment design effect violations by socioeconomic characteristic and list among Rajasthani women age 15 to 49[1]

	Design Effect P-Value	
	List 1	List 2
Age		
15-19	0.208	**0.001**
20-29	0.545	0.562
30-39	0.271	0.661
40-49	0.653	0.423
Marital status		
Currently married/cohabiting	0.645	0.567
Divorced or separated/widowed	0.523	1.000
Never married	0.450	0.079
School		
Never attended	**0.043**	0.998
Primary	0.557	0.222
Secondary	0.171	0.061
Higher or postgraduate	0.812	0.696
Wealth		
Poorest	0.318	0.260
Second poorest	0.912	0.509
Middle	0.473	0.428
Second wealthiest	0.063	0.701
Wealthiest	0.836	0.557
Caste of household head		
Scheduled caste	0.380	0.814
Scheduled tribe	0.527	**0.017**
Other backward caste	**0.002**	1.000
General	1.000	0.171
Religion of household head		
Hindu	0.574	**0.035**
Muslim	0.550	0.433
Other	0.281	0.289
Residence		
Rural	0.901	0.149
Urban	0.071	0.961
Parity		
0	0.501	**0.003**
1-2	0.596	0.219
3-4	0.125	1.000
5+	0.776	1.000
Abortion (direct question)		
No	0.513	**0.009**
Yes	0.320	0.996
Total	0.531	0.119

[1]Each list/subgroup specific p-value is Bonferroni-corrected to account for multiple comparison within the design effect test

Table 4.6a. List experiment estimate of lifetime experience of abortion using the piecewise estimator among Rajasthani women age 15 to 49 who reported abortion in direct question

		Number of reported items (proportion)						
List 1	Source	0	1	2	3	4	5	Sum
Row 1	List with abortion	0.000	0.152	0.454	0.331	0.050	0.014	1.000
Row 2	Proportion at least	1.000	1.000	0.848	0.394	0.063	0.014	--
Row 3	List without abortion	0.008	0.202	0.665	0.125	0.000	0.000	1.000
Row 4	Proportion at least	1.000	0.992	0.790	0.125	0.000	0.000	--
Row 5	Row 2 minus Row 4	0.000	0.008	0.058	0.269	0.063	0.014	0.411
Row 6	Exclude violations	0.000	0.008	0.058	0.269	0.063	0.014	0.411
List 2	Source							
Row 1	List with abortion	0.013	0.146	0.460	0.288	0.094	0.000	1.000
Row 2	Proportion at least	1.000	0.987	0.841	0.382	0.094	0.000	--
Row 3	List without abortion	0.091	0.550	0.268	0.091	0.000	0.000	1.000
Row 4	Proportion at least	1.000	0.909	0.359	0.091	0.000	0.000	--
Row 5	Row 2 minus Row 4	0.000	0.078	0.482	0.291	0.094	0.000	0.945
Row 6	Exclude violations	0.000	0.078	0.482	0.291	0.094	0.000	0.945
Average estimate across lists								67.79%
Average estimate across lists, violations excluded								67.79%

Note: Rows 1 and 3 represent the proportions reporting each number of items on the treatment and control lists, respectively. Rows 2 and 4 represent the proportions reporting *at least* each number of items on the treatment and control lists, respectively. Row 5 represents the difference between Row 2 and 4, which is equal to the proportion of women who report having an abortion and the total number of treatment list items indicated by the column (i.e. the joint proportion). Row 6 is a replicate of Row 5 where negative estimates have been excluded. The sum column for Row 5 and 6 represent the overall estimate of the proportion of women reporting a past abortion, including and excluding violations (i.e. negative joint proportions in Row 5), respectively.

Table 4.6b. List experiment estimate of lifetime experience of abortion using the piecewise estimator among Rajasthani women age 15 to 49 who did not report abortion in direct question

		Number of reported items (proportion)						
List 1	Source	0	1	2	3	4	5	Sum
Row 1	List with abortion	0.054	0.485	0.397	0.055	0.008	0.001	1.000
Row 2	Proportion at least	1.000	0.946	0.461	0.064	0.009	0.001	--
Row 3	List without abortion	0.043	0.480	0.420	0.052	0.004	0.000	1.000
Row 4	Proportion at least	1.000	0.957	0.477	0.056	0.004	0.000	--
Row 5	Row 2 minus Row 4	0.000	-0.010	-0.015	0.008	0.005	0.001	-0.011
Row 6	Exclude violations	0.000	0.000	0.000	0.008	0.005	0.001	0.014
List 2	Source							
Row 1	List with abortion	0.119	0.544	0.270	0.060	0.005	0.002	1.000
Row 2	Proportion at least	1.000	0.881	0.337	0.067	0.007	0.002	--
Row 3	List without abortion	0.120	0.503	0.294	0.077	0.005	0.000	1.000
Row 4	Proportion at least	1.000	0.880	0.377	0.082	0.005	0.000	--
Row 5	Row 2 minus Row 4	0.000	0.001	-0.039	-0.016	0.002	0.002	-0.051
Row 6	Exclude violations	0.000	0.001	0.000	0.000	0.002	0.002	0.004
Average estimate across lists								-3.08%
Average estimate across lists, violations excluded								0.95%

Note: Rows 1 and 3 represent the proportions reporting each number of items on the treatment and control lists, respectively. Rows 2 and 4 represent the proportions reporting *at least* each number of items on the treatment and control lists, respectively. Row 5 represents the difference between Row 2 and 4, which is equal to the proportion of women who report having an abortion and the total number of treatment list items indicated by the column (i.e. the joint proportion). Row 6 is a replicate of Row 5 where negative estimates have been excluded. The sum column for Row 5 and 6 represent the overall estimate of the proportion of women reporting a past abortion, including and excluding violations (i.e. negative joint proportions in Row 5), respectively.

5. Aim 2: Paradata as a Lens to Understand Underreporting of Abortion at the Individual Level

5.1 Background

Sensitive survey items are those that respondents perceive as intrusive and that raise respondent concerns regarding potential disclosure. These are items that could have legal or social consequences should information about respondent's behavior become known. As such, these survey items are prone to social desirability pressures (Tourangeau and Yan 2007). Including sensitive questions on a survey can affect the representativeness of responses as individuals decline to answer individual questions or decline to participate in the survey altogether (Tourangeau and Yan 2007). The validity of survey data on sensitive topics can also be poor as respondents may choose not to answer truthfully (Tourangeau and Yan 2007).

Questions on the topic of induced abortion experience are perceived as sensitive in nearly all societies. Survey-based research on abortion contends with bias to varying degrees through any or all of the aforementioned means. Prior research has been able to quantify the extent to which respondents answer questions about abortion truthfully by using clinical data as the gold standard. In the United States, as few as 35% to 48% of women reported known abortions in face-to-face interviews; women reported 73% of known abortions in Estonia where stigma is low (Anderson et al. 1994; Jones and Kost 2007).

The list experiment is an indirect technique developed to improve measurement of sensitive items on surveys by reducing the social desirability pressures of direct questioning. As described in Chapter 2, the standard list experiment randomizes individuals to either the treatment or control group. The control group is read a list of non-sensitive items, while the treatment group is read the same list, plus the sensitive item (Miller 1984). Interviewers then ask respondents to report *how many* of the items they have ever experienced, not *which* ones, without directly mentioning each item. Researchers can then calculate a simple difference in means between the total item counts of the treatment and control groups. The double list experiment is a modification that allows for every respondent to receive a treatment version of one list and a control version of another list, thus each respondent serves as control and treatment within the sample (Glynn 2013). The current work included a double list experiment to measure lifetime experience of induced abortion on a survey of reproductive age women in Rajasthan, India.

The survey experience is inherently social and requires the respondent and the interviewer to partake in an exchange ultimately governed by the same linguistic and social norms as a regular conversation (Sudman et al. 1996). The principles of conversations that also guide survey interactions expect that both speakers be truthful, contribute remarks germane to the conversation, make useful and new contributions to the interaction, and be clear (Grice 1975). These implicit expectations affect individuals' cognitive processes in preparing responses to survey questions through multiple means.

As outlined in Chapter 2, answering a survey question involves a number of cognitive stages. These stages and the specific pattern of failure in the context of a list experiment are described in Table 2.1 on page 27, which we reproduce below for reference (Sudman et al. 1996).

Table 2.1. Stages of list experiment responding and evidence of failures

Stage	Failure pattern	Null hypothesis
1) Decode the instructions	Groups with more schooling take less time to answer question regarding whether understand list experiment instructions	Equality of time spent listening to list experiment instructions and confirming understanding by schooling
	Groups with more schooling will take less time to answer the example list experiment question	Equality of time spent on list experiment example question by schooling
2) Interpret the question	No clear pattern; would need to conduct "think aloud" qualitative cognitive interview to determine how respondent interpreted the list experiment question	N/A
3) Retrieve and recall past life events	May intentionally skip or truncate this stage and simply estimate list experiment response, representing a "fake" or biased retrieval process	Equality of response time for people who express more social desirability pressure (not readily detectable from available data)
4) Enumerate the events	Groups who respond quickly on other questions that require similarly high levels of numeracy and cognitive ability will respond quickly to the list questions	No difference between time spent on similarly cognitively demanding questions and time spent on list experiment questions

5) Make a strategy regarding a response	Respondents who report an abortion on the direct question take longer to finalize their list experiment response if they considered editing their answer	Equality of time spent on treatment list experiment questions by direct abortion response

This chapter aims to analyze the active screen time paradata for the list experiment directions comprehension question, example question, and treatment and control list questions to assess response patterns that may provide information as to why the list experiment failed to produce higher estimates of abortion than direction questions in this context. With regard to question interpretation and understanding, we hypothesize that women with lower education will require the RE to explain the list experiment instructions repeatedly prior to confirming understanding and will take longer to answer the list experiment example question. With regard to response editing, we hypothesize that women who have had an abortion, as indicated by their response to the direct abortion questions, will take longer to respond to the treatment list as a result of delays caused by hesitation when deciding whether to include the abortion experience in their list response. We were unable to assess failures at other stages of the response process as the study was not designed to detect these issues.

5.2 Methodology

5.2.1 Data

Data for this aim come from Performance Monitoring and Accountability 2020 (PMA2020) data collection activities in Rajasthan, India (Zimmerman et al. 2017). The Indian Institute of Health Management and Research (IIHMR) conducted the data collection, with technical assistance provided by researchers from the Bill & Melinda Gates Institute for Reproductive Health at the Johns Hopkins Bloomberg School of Public Health. The sampling strategy was based on a probabilistic multi-stage cluster sampling design with probability proportional to size used to select enumeration areas (EA) with urban/rural strata and regions as the sampling domains. In Round 1 of data collection, resident enumerators (REs) mapped and listed the 147 selected EAs; the same 147 EA sampling frames were subsequently used in Round 2, which was when we included the list experiment questions. In Round 2, 35 households were randomly sampled from each EA. REs invited sampled households to participate in a brief household survey. REs then requested all eligible women, i.e. those age 15 to 49, to participate in an interview related to reproductive health. Prior to administering the survey, REs asked for consent for from all participants. The Institutional Review Boards (IRBs) at the Johns Hopkins Bloomberg School of Public Health and IIHMR provided ethical approval of the study protocol.

We included the list experiment and direct abortion questions in the female questionnaire. We randomized half of the respondents to receive treatment list A (i.e. including the sensitive item) along with control list B (i.e. not including the sensitive item). The other half of the respondents received control list A and treatment list B (Chapter 3, Table 3.1). In order to limit women's ability to determine the intent behind the list experiment questions, we placed them in the first section of the survey. We embedded the direct abortion questions in the reproductive history section.

In addition to the household and female survey data, paradata collected via log files on the smart phones used to conduct the surveys recorded the active screen time for each question. As such, the paradata provided the approximate response times for each survey question, enabling investigation of potential list experiment failures at certain stages of the response process. To examine whether REs characteristics impacted question response time, we also utilized the RE characteristics data, which came from a survey administered to REs after Round 2 data collection.

5.2.2 Analysis

We first conducted univariate analyses to examine the distribution of respondent and RE characteristics, as well as the distribution of response times for the relevant questions. We determined that the response times were all positively skewed due to long response time outliers, thus we generated logarithmic versions of the response time variables for subsequent multivariate regression analyses. These logarithmic versions of the variables were more normally distributed.

We then conducted bivariate analyses, employing adjusted Wald tests to investigate potential response time differences for abortion related questions by socioeconomic characteristics. We also re-calculated direct and list experiment estimates of lifetime experience of abortion using varying subgroups of responses based on response time percentile. We sought to assess whether potential RE misconduct or respondent disengagement resulting in suspiciously short response times may be a factor in the poor performance of the list experiment.

The first set of analyses examined possible failures in interpreting and understanding the question. To begin, we aimed to identify cognitively demanding questions to determine if there was a gradient in the response time by education level for questions unrelated to abortion. We used the birthdate question given the response time was among the longest of all the survey questions and RE feedback indicated it was often difficult for the respondent to provide both year and month. The independent variable, school, was coded as four indicator variables (never attended, primary, secondary, or higher or postgraduate school). We then studied the time required for the list experiment directions and associated yes/no question regarding instruction comprehension response time and the example list experiment question response time as the outcomes in successive analyses.

The second set of analyses checked for the presence of an editing process whereby women who have had an abortion take longer to respond to the list experiment treatment list question, which contained the "had an abortion" item. The exposure of interest was whether the respondent reported an abortion via the direct abortion questions. This variable was created using the initial yes/no question regarding any prior experience with a non-live birth, and the follow-up question regarding the outcome of the non-live birth (i.e. abortion, miscarriage, or stillbirth).

We modeled these multivariate analyses using ordinary least squares (OLS), fixed effects, and random effects; we used Hausman tests to determine which model was most appropriate given the observed data. We conducted these analyses using the non-logarithmic and the logarithmic versions of the response time variables. To isolate the within interviewer effects of respondent characteristics, we calculated cluster mean centered level-1 variables (Begg and Parides 2003). We also assessed the contribution of RE characteristics in the OLS and random effects models. For the final model, we added survey weights and calculated robust standard errors to account for clustering at the RE level. The OLS and random effects models tested for all analyses were of the form:

$$y_{ij} = \beta_0 + \beta_1^W(x_{1ij} - \bar{x}_{1i.}) + \beta_1^B \bar{x}_{1i.} + \boldsymbol{\beta_2^W}(\boldsymbol{X_{ij}} - \boldsymbol{\bar{X}_{i.}}) + \boldsymbol{\beta_3^B}\boldsymbol{\bar{X}_{i.}} + \boldsymbol{\beta_4^B}\boldsymbol{Z_j} + \varepsilon_i \qquad \text{(OLS)}$$

$$y_{ij} = \beta_0 + \beta_1^W(x_{1ij} - \bar{x}_{1i.}) + \beta_1^B \bar{x}_{1i.} + \boldsymbol{\beta_2^W}(\boldsymbol{X_{ij}} - \boldsymbol{\bar{X}_{i.}}) + \boldsymbol{\beta_3^B}\boldsymbol{\bar{X}_{i.}} + \boldsymbol{\beta_4^B}\boldsymbol{Z_j} + \varsigma_j \qquad \text{(Random)}$$

where y_{ij} is the response time for the outcome of interest (e.g. the list experiment directions comprehension question) for respondent *i* interviewed by RE *j*, β_0 is the overall response time average when all other variables are the reference category, x_{1ij} is the level-1 exposure variable of interest (e.g. education category), $\bar{x}_{1i.}$ is the cluster mean of the exposure variable, $x_{1ij} - \bar{x}_{1i.}$ is the cluster mean centered version of the exposure variable, β_1^W is the within interviewer effect of the exposure, β_1^B is the between interviewer effect of the exposure, X_{ij} is the vector of level-1 covariates, $\bar{X}_{i.}$ is the vector of cluster means level-1 covariates, $X_{ij} - \bar{X}_{i.}$ is the vector of cluster mean centered level-1 covariates, β_2^W is the vector of coefficients representing the within interviewer effect of the vector of level-1 covariates, β_3^B is a vector of coefficients representing the between interviewer effects of the vector of level-1 covariates, Z_j is the level-2 (interviewer) vector of covariates, β_4^B is the vector of coefficients representing the between interviewer effects of the vector of level-2 covariate, ς_j is the interviewer specific random effects, and ε_i is the error term for the OLS model. The random effects model assumes the expected value of the random intercepts is 0, and the random intercepts are assumed to be uncorrelated across interviewers and uncorrelated with level-1 residuals.

The fixed effects models we tested for all analyses were of the form:

$$y_{ij} - \bar{y}_{ij} = \beta_0 + \beta_1(x_{1ij} - \bar{x}_{1i.}) + \boldsymbol{\beta_2}(\boldsymbol{X_{ij}} - \boldsymbol{\bar{X}_{i.}}) + (\varepsilon_{ij} - \bar{\varepsilon}_{i.}) \text{ (Fixed)}$$

where y_{ij} is the response time for the outcome of interest for respondent i interviewed by RE j, $\bar{y}_{i.}$ is the cluster mean response time for the outcome, $y_{ij} - \bar{y}_{i.}$ is the cluster mean centered response time for the outcome, β_0 is the overall response time average when all other variables are the reference category, x_{1ij} is the level-1 exposure variable of interest, $\bar{x}_{1i.}$ is the cluster mean of the level-1 exposure variable, $x_{1ij} - \bar{x}_{1i.}$ is the cluster mean centered level-1 exposure, β_1 represents the within interviewer effect of the exposure variable on the response time, X_{ij} is the vector of level-1 covariates, $\bar{X}_{i.}$ is the vector of cluster mean level-1 covariates, $X_{ij} - \bar{X}_{i.}$ is the vector of cluster mean centered level-1 covariates, β_2 is the vector of coefficients representing the within interviewer effects of the vector of level-1 covariates, ε_{ij} is the error term, $\bar{\varepsilon}_{i.}$ is the cluster mean error term, and $\varepsilon_{ij} - \bar{\varepsilon}_{i.}$ is the cluster mean centered error term, where the expectation of the error is 0, and the interviewer specific fixed effects parameters are unknown and fixed.

Respondent level covariates included age, marital status, education, wealth quintile, caste, religion, residence, parity, and RE-respondent acquaintance indicator variables, along with a continuous list experiment control list question response time variable. RE level covariates investigated included age, education, whether ever married, whether the RE thought abortion was legal under any circumstances, and whether the RE thought the list question was difficult to implement. We conducted the analyses in Stata version 15 and assessed statistical significant using an alpha of 0.05 (StataCorp 2017).

5.3 Results

In total, 6,017 women age 15 to 49 from selected households completed the female survey with corresponding paradata. For a description of the sample characteristics, refer to Table 4.1 in Chapter 4. The response rate overall was 97.8%, and the response rates for the initial direct abortion question (regarding past non-live birth), the question about how the non-live birth ended, and the list experiment treatment list (with the abortion item), were all 99.9%. Reported lifetime experience of abortion was 3.5% (standard error (SE) 0.7) via the direct questions and 1.8% (SE 0.7) via the list experiment (results not shown).

Examining response times for the direct question regarding past experience with a non-live birth, women took on average 11.0 seconds (SE 0.4) to respond (Table 5.1). The list experiment related questions required much longer response times; on average 24.4 (SE 1.5), 23.3 (SE 1.2), 27.0 (SE 1.4), and 23.7 (SE 1.3) for the list experiment directions and associated "yes" or "no" comprehension question, the list experiment example question including food items, the control list experiment question, and the treatment list experiment question (Table 5.1). Adjusted Wald test results revealed significant variation in response times for several questions within a number of socioeconomic characteristics. Older women consistently took longer to respond to abortion and list experiment questions, as did women with more past births and who reported an abortion on the direct questions. In contrast, women who had never married tended to respond more quickly than women with other marital statuses.

Birthdate, which we used as our non-sensitive but cognitively demanding reference, took women 52.4 seconds (SE 1.9) to answer; this was much longer than any of the list experiment or direct abortion questions (Table 5.1). We detected significant variation in birthdate response time by age, marital status, and parity in a similar manner to the list experiment questions. However, response times by school and caste were also statistically significantly different. Additionally, birthdate response time was *not* significantly different by direct reporting of abortion.

To investigate the potential role of comprehension delays in explaining the list experiment's failure to produce more valid estimates of abortion, we first sought to establish a relationship between education and birthdate – a similarly cognitively demanding question. The bivariate analyses revealed a highly statistically significant association between birthdate response time and education, whereby increasing levels of education were associated with decreasing response times. Response times for never attended, attended primary school, attended secondary school, and attended higher education were 56.0, 53.7, 52.2, and 44.7 seconds, respectively (p-value<0.001) (Table 5.1). In the multivariate analysis, we systematically added the independent variable, the respondent characteristics, and then the RE characteristics (in the random effects model). The Hausman test showed that the random effects model was the preferred specification. In the preferred random effects model, women with higher education responded to the birthdate question 6.16 (95% confidence interval (CI) -10.50, -1.81) seconds faster than women who had never attended school; response times for women who attended primary or secondary school did not differ significantly (Appendix 5.1).

To determine whether a similar comprehension delay occurred for the list experiment questions, we studied the time required to answer the list experiment directions comprehension yes/no question, as well as the list experiment example question. The bivariate results revealed no significant variability in the list experiment directions comprehension question or the list experiment example question response times by education (Table 5.2). Hausman tests preferred the fixed effects specification compared to the random effects models. The final model revealed no significant association between any level of education and response time for the list experiment directions comprehension (Table 5.2) and list experiment example questions (Table 5.3). The final models examining the role of education on the list experiment directions comprehension question (Table 5.2) response time and the list experiment example question response time (Table 5.3) had large rhos of 0.39 and 0.21, respectively, suggesting the RE behavior had a significant impact on response times for these questions.

Results revealed editing during responding to the list experiment treatment list as a potential explanation of its failure to produce more valid estimates of abortion. The bivariate results detected a statistically significant association, whereby women who reported a past abortion via the direct questions took 45.1 (SE 4.4) seconds to respond and women who reported no abortion took 22.9 (SE 1.2) seconds to respond ($p<0.001$) (Table 5.1). In the most preferred fixed effects model, women who reported an abortion on the direct abortion questions took 11.6 (95% CI 7.2, 16.0) seconds longer to respond to the list experiment treatment list compared to women who did not report an abortion on the direct abortion questions, adjusting for respondent characteristics and RE effects. Greater parity was significantly associated with longer treatment list response times, with women who had 1 to 2 children (2.1; 95% CI 0.6, 3.7), 3 to 4 children (3.8; 95% CI 1.9, 5.7), or 5 or more children (5.6; 95% CI 2.6, 8.6) taking significantly longer to respond compared to nulliparous women (Table 5.4). In contrast, being never married (as opposed to currently married or cohabiting) was associated with a significantly shorter response time (-3.2; 95% CI -5.1, -1.3), as was being of a scheduled tribe (as opposed to a scheduled caste) (-2.9; 95% CI -4.9, -0.9). Again, the rho in the final model was fairly large at 0.18, indicating that RE behavior impacted response times for the treatment list question.

The robustness of the results regarding editing issues was tested by examining whether abortion was associated with the response times for a number of other questions. These sensitivity analyses included testing the same fixed effects model on the time required to respond to the list experiment directions comprehension question, list experiment example question, list experiment control list question, birthdate question, education question, marital status question, and ever non-live birth question. Only the ever non-live birth question was associated with a borderline statistically significant association with whether a woman reported an abortion via the direction questions; none of the other questions revealed a statistically significant association between reporting an abortion via the direct questions and response time (results not shown). We also conducted the multivariate regression models using the logarithmic version of the response time variables and the results were qualitatively the same. For ease of interpretation we presented only the non-logarithmic results.

Lastly, we examined the potential role of RE misconduct or respondents' disengagement with regard to the list experiment results. REs hurrying through the questionnaire and/or falsifying data could explain very short response times for list experiment questions. This RE misbehavior may have partially contributed to the high rhos in the abovementioned final models. But respondents failing to engage in the retrieval process could also explain the short response times. Either explanation would result in invalid data and poor list experiment performance. We estimated overall direct and list experiment lifetime experience of abortion among subsets of the sample based on percentile of the response time to: (1) the list experiment directions comprehension question, (2) the list experiment example question, and (3) the control and treatment list experiment questions. With regard to the list experiment directions comprehension question, estimates of abortion prevalence were fairly stable via the direct questions across low and high percentiles of response times. This pattern was also true of the list experiment estimates, with the exception of the analysis using only those whose response times were above the 95th percentile; among respondents in this sub-sample, the lifetime experience of abortion estimate via the list experiment questions was 6.4% (Figure 5.1a). The estimates of lifetime abortion experience from the direct and list experiment questions both had a small increasing gradient as the list experiment example question response time increased (Figure 5.1b). Among subsets of the population with increasingly longer response times to both the control and treatment list experiment questions, estimates of abortion experience were also increasingly higher (Figure 5.1c). Most notably, using only the data for those whose response times to the list experiment control and treatment list questions were above the 95th percentile resulted in a direct abortion estimate of 9.4% but a list experiment abortion estimate of 20.1% (Figure 5.1c).

5.4 Discussion

Findings do not support the hypothesis that cognitive issues in interpreting the list experiment directions and questions contributed to the failure of the list experiment to produce more valid estimates of abortion than the direct questions. However, there is evidence of editing delays for women who have experienced an abortion. The delay may be triggered by confrontation with a list that includes the word "abortion". It could be an affective process of recalled emotion and/or a process of rationally choosing whether to edit a response for social desirability. Since underreporting on direct abortion questions is substantial (Jones and Kost 2007), the results are likely conservative given that women who hesitated in providing their response and ultimately responded "no" on the direct abortion question would be in the reference group and thus bias the effect estimate towards the null. Additionally, findings using subsets of the sample based on the list experiment question response time percentiles suggest a possible role of RE misconduct and/or respondent disengagement as a potential explanation for the poor performance of the list experiment. The bias that can arise as a result of low overall or individual question response rates does not appear to explain the low abortion estimates.

The finding that women who had experienced a prior abortion took significantly longer to respond to a list experiment question that included the abortion item adds support to the idea that respondents engage in an editing process when asked about sensitive items on surveys. Prior research has demonstrated longer response times on survey questions that involve reporting socially undesirable behaviors or ideas (Holtgraves 2004; Holtgraves et al. 1997).

This study is the first investigation the authors know of that has applied social and cognitive psychology theories in an effort to better understand the mechanism by which abortion is underreported on surveys. This analysis presents new methods others can consider and continue to develop, leveraging the paradata that is increasingly collected in the course of survey implementation (Kreuter et al. 2010). The quality of the response time data is precise because it is recorded by the smart phones used for the data collection.

Although these indirect means of evaluating sources of bias in reporting sensitive items can be informative, the data collection activities were not explicitly designed to assess these hypotheses. As such, these findings offer only initial evidence regarding this response phenomenon, but alternative explanations cannot be ignored. The cognitive ability and numeracy required to respond to the list experiment questions using only a numeric response, thus maintaining the confidentiality afforded by the list experiment design, may not be adequately quantified by the education indicator variables. Response time is at best a proxy for capturing the interviewer-respondent interaction, but other occurrences during the survey could result in longer response times. Additionally, to the extent that delays are significant, the phone goes into an energy saving mode and this time is not captured in the response time paradata. If these longer delays occurred systematically for certain types of women, this may have introduced bias.

Further research leveraging paradata from smart phone data collection could improve our understanding of the cognitive processes that respondents experience when answering sensitive survey questions. In addition, more use of qualitative cognitive interviews would inform our knowledge of respondent interpretation of abortion-related question wording and recall and retrieval. Existing evidence suggests abortion is subject to multiple understandings and terms are prone to misinterpretation. Qualitative cognitive interviews would enable assessment of respondents' ability to retrieve and enumerate past events in the case of the list experiment. This would help improve the design of future abortion measurement studies.

Table 5.1. Average response time in seconds among Rajasthani women age 15 to 49, by socioeconomic characteristics and question (N=6,017)[1]

	Ever non-live birth	List directions	Example list	Control list	Treatment list	Birthdate
	Time (SE)					
Age						
15-19	8.6 (0.5)***	22.9 (1.7)	20.4 (1.3)***	20.1 (1.3)***	16.9 (1.2)***	42.6 (1.9)***
20-29	11.2 (0.5)***	24.7 (1.7)	22.9 (1.3)***	27.4 (1.5)***	24.3 (1.5)***	49.8 (2.1)***
30-39	11.8 (0.5)***	25.2 (1.7)	24.9 (1.4)***	29.4 (1.8)***	26.7 (1.6)***	56.3 (2.4)***
40-49	12.0 (0.6)***	24.3 (1.8)	24.6 (1.6)***	29.9 (1.9)***	24.9 (1.6)***	61.8 (2.8)***
Marital status						
Currently married/cohabiting	11.8 (0.4)***	24.7 (1.6)	24.0 (1.2)***	29.1 (1.5)***	25.8 (1.4)***	54.9 (2.1)***
Divorced or separated/widowed	13.8 (1.5)***	26.3 (3.3)	26.9 (3.0)***	30.2 (3.7)***	27.3 (3.2)***	61.7 (6.0)***
Never married	8.1 (0.5)***	23.3 (1.6)	20.3 (1.3)***	19.3 (1.2)***	15.9 (1.1)***	42.2 (1.9)***
School						
Never attended	11.4 (0.5)	24.1 (1.6)	23.5 (1.2)	28.2 (1.6)	25.1 (1.5)	56.0 (2.5)***
Primary	11.5 (0.6)	24.1 (1.9)	23.9 (1.6)	28.3 (1.9)	24.5 (1.6)	53.7 (2.6)***
Secondary	10.4 (0.6)	24.5 (2.1)	21.9 (1.5)	25.5 (1.8)	21.8 (1.6)	52.2 (2.5)***
Higher or postgraduate	10.4 (0.5)	25.3 (2.3)	23.3 (1.7)	24.7 (1.7)	21.8 (1.5)	44.7 (2.2)***
Wealth						
Poorest	10.5 (0.5)	21.0 (2.0)	21.0 (1.5)	23.8 (2.0)	20.8 (1.6)	53.0 (3.1)
Second poorest	11.6 (0.6)	24.8 (2.2)	24.0 (1.4)	27.6 (1.8)	24.5 (1.6)	52.8 (3.4)
Middle	11.2 (0.6)	25.5 (1.9)	24.1 (1.5)	29.5 (2.0)	24.9 (1.8)	52.2 (2.7)
Second wealthiest	11.2 (0.6)	25.3 (1.9)	22.9 (1.7)	27.9 (2.0)	24.4 (1.8)	50.6 (2.0)
Wealthiest	10.6 (0.6)	24.8 (2.2)	23.9 (2.0)	26.0 (1.9)	23.4 (1.9)	53.2 (3.1)
Caste of household head						
Scheduled caste	10.8 (0.6)	21.1 (2.2)*	22.8 (1.7)	25.6 (2.1)	23.4 (2.0)	47.2 (2.6)**
Scheduled tribe	10.6 (0.5)	20.7 (2.4)*	21.2 (1.7)	24.4 (2.2)	20.4 (2.2)	50.4 (2.9)**
Other backward caste	11.2 (0.6)	26.8 (2.2)*	24.1 (1.7)	28.6 (1.8)	24.8 (1.7)	53.5 (2.9)**
General	11.3 (0.6)	26.5 (2.6)*	23.9 (1.7)	27.7 (2.0)	24.6 (1.8)	57.3 (3.2)**
Religion of household head						
Hindu	11.1 (0.4)	24.2 (1.7)	23.6 (1.3)*	26.7 (1.4)	23.5 (1.3)	52.6 (2.2)

Muslim	10.4 (0.7)	25.7 (3.8)	20.4 (1.9)*	28.9 (3.7)	24.5 (2.8)	50.4 (2.7)
Other	10.3 (1.3)	24.8 (3.7)	29.0 (3.9)*	28.8 (5.4)	27.4 (5.2)	54.1 (5.2)
Residence						
Rural	11.3 (0.5)	25.0 (2.0)	23.5 (1.3)	27.3 (1.5)	23.7 (1.4)	52.1 (2.5)
Urban	10.5 (0.7)	23.4 (2.5)	22.9 (2.2)	26.5 (2.8)	23.6 (2.5)	52.8 (3.0)
Parity						
0	9.0 (0.4)***	22.8 (1.5)**	20.4 (1.1)***	20.4 (1.2)***	17.1 (1.1)***	43.1 (1.8)***
1-2	11.7 (0.5)***	25.0 (1.9)**	23.6 (1.6)***	28.2 (1.7)***	24.9 (1.6)***	54.5 (2.3)***
3-4	11.8 (0.5)***	25.1 (1.6)**	24.8 (1.2)***	31.0 (1.7)***	27.8 (1.7)***	59.1 (2.7)***
5+	13.4 (0.8)***	26.1 (2.0)**	28.3 (1.9)***	34.9 (2.3)***	31.0 (1.8)***	57.8 (4.1)***
Abortion (direct question)						
No	10.9 (0.4)	24.2 (1.6)*	23.0 (1.1)***	26.5 (1.4)***	22.9 (1.2)***	52.2 (1.9)
Yes	13.2 (1.4)	29.1 (2.8)*	31.6 (2.6)***	40.8 (3.2)***	45.1 (4.4)***	57.2 (4.9)
Total	11.0 (0.4)	24.4 (1.5)	23.3 (1.2)	27.0 (1.4)	23.7 (1.3)	52.4 (1.9)

[1]All estimates include weights accounting for complex survey design and non-response
* denotes p<0.10, ** denotes p<0.05, and *** denotes p<0.01; based on adjusted Wald test of null hypothesis that all response times were equal within the category

Table 5.2. Comparison of multivariate linear regression models of *education* on *list experiment directions question* response time employing ordinary least squares (OLS), as well as fixed effects (FE) and random effects (RE) that account for clustering at the interviewer level

	OLS		FE		RE		FE + Robust SE and weights	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Schooling (reference never attended)								
Primary school	0.38	-1.24,1.99	0.37	-0.99,1.73	0.37	-0.99,1.73	0.34	-1.05,1.73
Secondary school	0.06	-1.95,2.07	0.01	-1.69,1.70	0.01	-1.69,1.70	0.63	-1.33,2.59
Higher education	0.49	-1.68,2.65	0.47	-1.36,2.30	0.47	-1.36,2.30	1.10	-1.01,3.22
Control list response time	0.21***	0.19,0.24	0.21***	0.19,0.24	0.21***	0.19,0.24	0.21***	0.18,0.25
Age (reference 15-19)								
20-29	1.26	-1.00,3.53	1.39	-0.53,3.31	1.38	-0.53,3.30	1.24	-0.85,3.33
30-39	2.13	-0.58,4.84	2.27	-0.02,4.56	2.26	-0.03,4.55	1.64	-0.67,3.96
40-49	0.87	-2.11,3.85	0.98	-1.54,3.50	0.97	-1.55,3.49	0.37	-2.11,2.84
Marital status (reference currently married/cohabiting)								
Divorced/widowed	0.81	-2.66,4.28	0.81	-2.13,3.74	0.81	-2.13,3.74	1.86	-1.89,5.61
Never married	0.97	-1.60,3.55	1.05	-1.13,3.22	1.04	-1.13,3.22	0.73	-1.44,2.90
Wealth quintile (reference poorest)								
Middle poorest	-0.07	-2.24,2.11	0.00	-1.84,1.83	0.00	-1.84,1.83	0.47	-1.24,2.18
Middle	1.24	-1.06,3.54	1.22	-0.72,3.17	1.22	-0.72,3.17	1.26	-0.60,3.12
Middle wealthiest	1.54	-0.93,4.00	1.52	-0.56,3.61	1.53	-0.56,3.61	1.86	-0.17,3.90
Wealthiest	0.56	-2.13,3.26	0.54	-1.73,2.82	0.54	-1.73,2.82	0.78	-1.43,2.99
Caste (reference scheduled caste)								
Scheduled tribe	-0.13	-2.56,2.29	-0.09	-2.14,1.96	-0.09	-2.14,1.96	0.34	-2.14,2.83
Other backward caste	-0.03	-1.95,1.88	0.01	-1.60,1.63	0.01	-1.61,1.63	0.53	-1.55,2.60
Generate caste	0.82	-1.46,3.11	0.86	-1.07,2.79	0.86	-1.07,2.79	1.51	-0.88,3.90
Religion (reference Hindu)								
Muslim	-1.34	-4.22,1.55	-1.39	-3.82,1.05	-1.38	-3.82,1.05	0.03	-4.05,4.11
Other religion	3.03	-2.34,8.40	3.03	-1.51,7.56	3.03	-1.51,7.56	4.28	-1.20,9.76
Residence (reference rural)								

Urban	-89.46	-341.73,162.81	na	na	-76.40	-880.60,727.79	na	na
Parity (reference 0)								
1-2	-0.63	-2.81,1.55	-0.60	-2.44,1.24	-0.60	-2.44,1.24	-0.32	-2.43,1.80
3-4	-1.24	-3.68,1.20	-1.21	-3.27,0.84	-1.21	-3.27,0.85	-0.87	-3.10,1.36
5+	-1.65	-4.73,1.42	-1.64	-4.23,0.96	-1.64	-4.23,0.96	-1.03	-3.91,1.86
Interviewer acquainted (reference no)								
Yes	0.37	-1.66,2.40	0.38	-1.33,2.10	0.38	-1.33,2.10	1.1	-1.98,4.18
Constant	-45.81***	-59.38,-32.23	17.07***	14.05,20.08	-44.79	-94.35,4.78	15.32***	11.85,18.80
Rho	na		0.38		0.32		0.39	
AIC	54234.94		52025.28		na		51819.47	
N	6,017		6,017		6,017		6,017	

* denotes p<0.10, ** denotes p<0.05, and *** denotes p<0.01

Table 5.3. Comparison of multivariate linear regression models of *education* on *list experiment example question* response time employing ordinary least squares (OLS), as well as fixed effects (FE) and random effects (RE) that account for clustering at the interviewer level

	OLS		FE		RE		FE + Robust SE and weights	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Schooling (reference never attended)								
Primary school	0.01	-1.42,1.45	-0.01	-1.36,1.35	-0.01	-1.36,1.35	0.29	-1.35,1.92
Secondary school	-1.56	-3.35,0.22	-1.57	-3.26,0.12	-1.57	-3.26,0.12	-1.71	-3.51,0.09
Higher education	-0.82	-2.74,1.10	-0.84	-2.66,0.98	-0.84	-2.66,0.98	-1.36	-3.94,1.22
Control list response time	0.37***	0.35,0.39	0.37***	0.35,0.39	0.37***	0.35,0.39	0.36***	0.31,0.40
Age (reference 15-19)								
20-29	-0.22	-2.24,1.79	-0.27	-2.18,1.64	-0.26	-2.17,1.65	-0.15	-2.23,1.92
30-39	-0.36	-2.77,2.05	-0.41	-2.69,1.87	-0.40	-2.68,1.88	0.41	-2.10,2.91
40-49	-1.68	-4.33,0.97	-1.74	-4.25,0.77	-1.73	-4.24,0.77	-0.54	-3.56,2.48
Marital status (reference currently married/cohabiting)								
Divorced/widowed	0.95	-2.14,4.03	0.95	-1.97,3.86	0.95	-1.97,3.86	0.82	-2.35,4.00
Never married	-0.83	-3.12,1.46	-0.86	-3.03,1.30	-0.86	-3.02,1.31	-0.43	-2.72,1.86
Wealth quintile (reference poorest)								
Middle poorest	1.93	-0.00,3.86	1.94*	0.12,3.77	1.94*	0.12,3.77	1.72	-0.63,4.07
Middle	1.64	-0.40,3.69	1.66	-0.27,3.59	1.66	-0.28,3.59	1.14	-0.97,3.26
Middle wealthiest	1.85	-0.34,4.04	1.86	-0.21,3.93	1.86	-0.22,3.93	1.28	-1.31,3.87
Wealthiest	0.90	-1.49,3.30	0.92	-1.35,3.18	0.92	-1.35,3.18	0.93	-1.49,3.34
Caste (reference scheduled caste)								
Scheduled tribe	0.79	-1.37,2.95	0.78	-1.26,2.82	0.78	-1.26,2.82	0.45	-1.77,2.66
Other backward caste	-0.93	-2.63,0.77	-0.95	-2.56,0.67	-0.94	-2.55,0.67	-1.20	-3.35,0.95
Generate caste	-0.06	-2.09,1.97	-0.06	-1.98,1.86	-0.06	-1.98,1.86	-0.03	-2.29,2.23
Religion (reference Hindu)								
Muslim	-1.34	-3.90,1.23	-1.33	-3.75,1.09	-1.33	-3.75,1.09	-2.26	-5.76,1.23
Other religion	2.95	-1.82,7.73	2.95	-1.56,7.47	2.95	-1.56,7.47	2.17	-2.17,6.51

Residence (reference rural)								
Urban	-67.43	-291.66,156.80	na	na	-57.64	-534.22,418.94	na	na
Parity (reference 0)								
1-2	-1.05	-2.99,0.88	-1.05	-2.88,0.78	-1.05	-2.88,0.78	-1.37	-3.50,0.76
3-4	-0.46	-2.63,1.70	-0.45	-2.50,1.60	-0.45	-2.50,1.59	-1.01	-3.48,1.45
5+	1.53	-1.20,4.26	1.53	-1.06,4.11	1.53	-1.06,4.11	0.83	-2.31,3.96
Interviewer acquainted (reference no)								
Yes	1.01	-0.80,2.81	1.01	-0.70,2.72	1.01	-0.70,2.71	1.58	-1.35,4.50
Constant	-20.23**	-32.29,-8.16	13.61***	10.61,16.61	-20.26	-49.20,8.67	13.61***	9.46,17.77
Rho	na		0.20		0.12		0.21	
AIC	52816.86		51972.76		na		51975.75	
N	6,017		6,017		6,017		6,017	

* denotes p<0.10, ** denotes p<0.05, and *** denotes p<0.01

Table 5.4. Comparison of multivariate linear regression models of *direct abortion reporting* on *list experiment treatment list* response time employing ordinary least squares (OLS), as well as fixed effects (FE) and random effects (RE) that account for clustering at the interviewer level

	OLS		FE		RE		FE + Robust SE and weights	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Direct abortion question response (reference no)								
Yes	9.10***	6.26,11.95	9.10***	6.29,11.91	9.10***	6.29,11.91	11.59***	7.15,16.03
Control list response time	0.44***	0.42,0.46	0.44***	0.42,0.46	0.44***	0.42,0.46	0.41***	0.36,0.46
Age (reference 15-19)								
20-29	0.24	-1.60,2.08	0.2	-1.62,2.02	0.22	-1.60,2.04	-0.11	-1.65,1.43
30-39	-0.25	-2.46,1.95	-0.28	-2.46,1.89	-0.27	-2.44,1.91	-0.62	-2.57,1.32
40-49	-2.23	-4.65,0.20	-2.26	-4.65,0.14	-2.24	-4.63,0.15	-2.43*	-4.64,-0.23
Marital status (reference currently married/cohabiting)								
Divorced/widowed	0.32	-2.50,3.14	0.32	-2.46,3.11	0.32	-2.46,3.11	-0.16	-4.74,4.43
Never married	-3.19**	-5.29,-1.10	-3.21**	-5.27,-1.14	-3.20**	-5.26,-1.13	-3.22**	-5.16,-1.29
Schooling (reference never attended)								
Primary school	0.78	-0.53,2.09	0.78	-0.52,2.07	0.78	-0.51,2.07	0.23	-1.39,1.86
Secondary school	1.01	-0.62,2.64	1.00	-0.61,2.62	1.01	-0.60,2.62	-0.03	-2.44,2.38
Higher education	0.86	-0.90,2.62	0.86	-0.88,2.59	0.86	-0.88,2.59	0.1	-2.51,2.72
Wealth quintile (reference poorest)								
Middle poorest	1.56	-0.21,3.32	1.56	-0.18,3.30	1.56	-0.18,3.30	1.73	-0.39,3.86
Middle	1.75	-0.12,3.62	1.76	-0.09,3.60	1.75	-0.09,3.60	1.47	-0.64,3.58
Middle wealthiest	2.25*	0.24,4.25	2.26*	0.28,4.24	2.25*	0.28,4.23	1.86	-0.63,4.35
Wealthiest	0.81	-1.38,3.00	0.82	-1.34,2.98	0.82	-1.34,2.98	1.35	-1.28,3.98
Caste (reference scheduled caste)								
Scheduled tribe	-2.93**	-4.91,-0.96	-2.94**	-4.89,-0.99	-2.94**	-4.88,-0.99	-3.60**	-5.77,-1.43
Other backward caste	-2.42**	-3.98,-0.87	-2.43**	-3.96,-0.89	-2.43**	-3.96,-0.89	-2.22	-4.64,0.21
Generate caste	0.31	-1.54,2.17	0.31	-1.52,2.14	0.31	-1.52,2.14	0.31	-2.46,3.07
Religion (reference Hindu)								

Muslim	-0.86	-3.20,1.48	-0.87	-3.18,1.44	-0.86	-3.17,1.45	-0.37	-2.85,2.10
Other religion	-2.11	-6.47,2.26	-2.11	-6.42,2.20	-2.11	-6.42,2.20	-1.39	-6.31,3.52
Residence (reference rural)								
Urban	144.89	-61.16,350.94	na	na	137.00	-153.82,427.81	na	na
Parity (reference 0)								
1-2	2.06*	0.29,3.84	2.08*	0.32,3.83	2.07*	0.32,3.82	2.13**	0.55,3.70
3-4	4.02***	2.03,6.00	4.03***	2.07,5.99	4.02***	2.06,5.98	3.82***	1.95,5.70
5+	5.17***	2.67,7.67	5.18***	2.71,7.65	5.18***	2.71,7.64	5.59***	2.59,8.59
Interviewer acquainted (reference no)								
Yes	1.50	-0.15,3.15	1.50	-0.13,3.13	1.50	-0.13,3.13	1.97	-0.06,4.00
Constant	-2.00	-13.07,9.07	9.42***	6.56,12.29	-4.04	-20.98,12.91	10.21***	6.50,13.91
Rho	na		0.17		0.03		0.18	
AIC	51740.7		51412.93		na		51499.32	
N	6,017		6,017		6,017		6,017	

* denotes p<0.10, ** denotes p<0.05, and *** denotes p<0.01

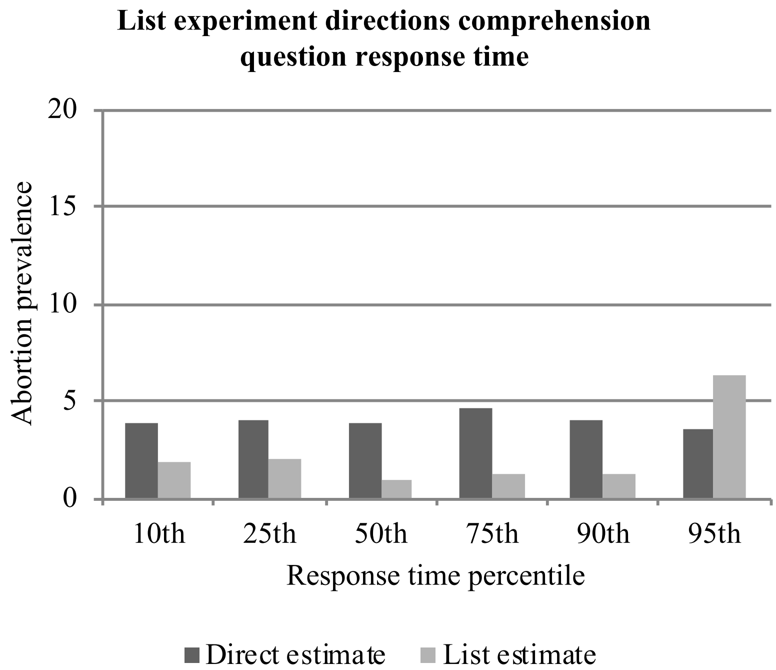
Appendix 5.1. Comparison of multivariate linear regression models of *education* on *birthdate* response time employing ordinary least squares (OLS), as well as fixed effects (FE) and random effects (RE) that account for clustering at the interviewer level

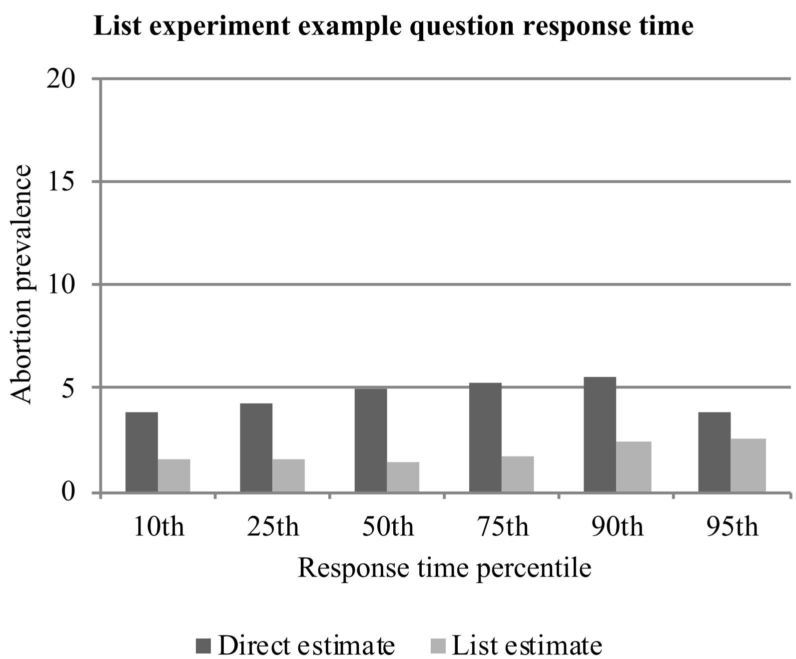
	OLS		FE		RE		RE + Robust SE	
	β	95% CI	β	95% CI	β	95% CI	β	95% CI
Schooling (reference never attended)								
Primary school	-1.40	-4.35,1.54	-1.48	-4.19,1.22	-1.48	-4.18,1.23	-1.48	-4.35,1.40
Secondary school	-0.46	-4.13,3.21	-0.54	-3.91,2.83	-0.53	-3.90,2.84	-0.53	-4.02,2.96
Higher education	-5.93**	-9.89,-1.97	-6.00**	-9.63,-2.37	-5.99**	-9.62,-2.36	-5.99**	-9.68,-2.30
Control list response time	0.27***	0.22,0.31	0.27***	0.22,0.31	0.27***	0.22,0.31	0.27***	0.19,0.34
Age (reference 15-19)								
20-29	1.62	-2.53,5.76	1.74	-2.07,5.55	1.73	-2.08,5.54	1.73	-1.50,4.95
30-39	5.20*	0.24,10.16	5.27*	0.72,9.82	5.26*	0.71,9.81	5.26*	1.11,9.41
40-49	11.88***	6.43,17.34	11.91***	6.91,16.92	11.91***	6.90,16.91	11.91***	6.56,17.25
Marital status (reference currently married/cohabiting)								
Divorced/widowed	-1.89	-8.24,4.46	-1.91	-7.74,3.91	-1.91	-7.74,3.91	-1.91	-8.00,4.17
Never married	-0.38	-5.09,4.33	-0.27	-4.59,4.05	-0.28	-4.60,4.04	-0.28	-3.79,3.23
Wealth quintile (reference poorest)								
Middle poorest	-1.43	-5.40,2.54	-1.30	-4.95,2.34	-1.32	-4.96,2.33	-1.32	-4.88,2.25
Middle	-1.75	-5.96,2.45	-1.79	-5.65,2.07	-1.79	-5.65,2.07	-1.79	-5.75,2.17
Middle wealthiest	-0.45	-4.96,4.06	-0.49	-4.63,3.64	-0.49	-4.63,3.65	-0.49	-5.03,4.05
Wealthiest	-0.48	-5.40,4.45	-0.50	-5.02,4.02	-0.50	-5.02,4.02	-0.50	-5.43,4.43
Caste (reference scheduled caste)								
Scheduled tribe	-1.64	-6.08,2.80	-1.65	-5.72,2.43	-1.65	-5.72,2.43	-1.65	-6.07,2.77
Other backward caste	1.08	-2.42,4.59	1.08	-2.14,4.29	1.08	-2.14,4.29	1.08	-2.44,4.60
Generate caste	2.81	-1.36,6.99	2.80	-1.03,6.63	2.80	-1.03,6.63	2.80	-1.64,7.25
Religion (reference Hindu)								
Muslim	-1.93	-7.20,3.34	-1.94	-6.77,2.90	-1.94	-6.77,2.90	-1.94	-6.61,2.74
Other religion	-2.56	-12.38,7.26	-2.55	-11.56,6.46	-2.55	-11.56,6.46	-2.55	-14.02,8.91
Residence (reference rural)								
Urban	424.11	-37.21,885.43	na	na	417.14	-728.90,1563.18	417.14	-83.47,917.74

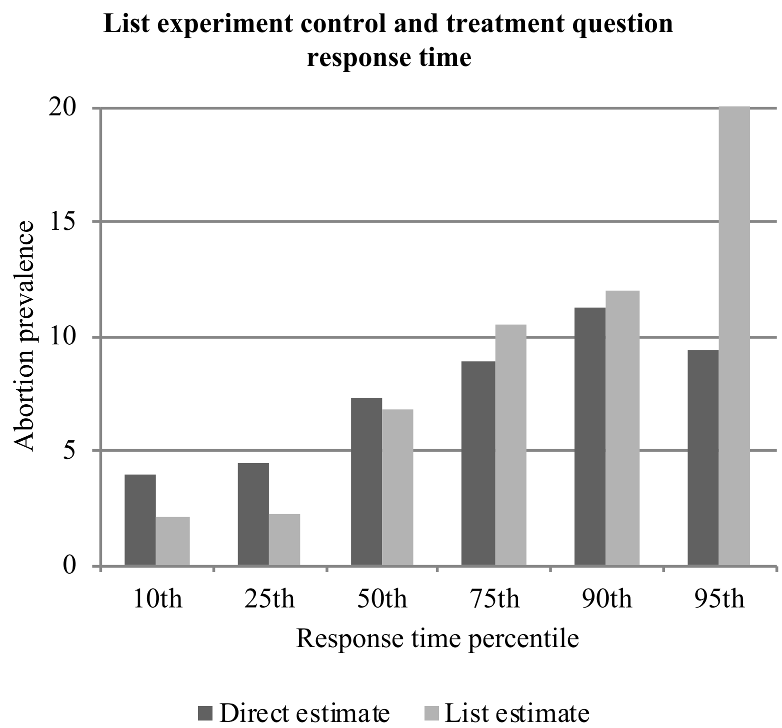
Parity (reference 0)								
1-2	3.75	-0.23,7.74	3.84*	0.18,7.50	3.83*	0.17,7.49	3.83*	0.51,7.15
3-4	4.15	-0.31,8.60	4.25*	0.16,8.34	4.24*	0.15,8.33	4.24*	0.20,8.29
5+	2.17	-3.45,7.79	2.21	-2.95,7.36	2.21	-2.95,7.36	2.21	-3.57,7.98
Interviewer acquainted (reference no)								
Yes	2.31	-1.40,6.03	2.35	-1.06,5.76	2.35	-1.06,5.76	2.35	-1.60,6.30
Constant	58.96***	34.14,83.78	38.33***	32.34,44.33	58.23	-11.86,128.32	58.23	-5.96,122.42
Rho	na		0.24		0.18		0.18	
AIC	61498.36		60289.95		na		na	
N	6,017		6,017		6,017		6,017	

* denotes p<0.10, ** denotes p<0.05, and *** denotes p<0.01

Figure 5.1. Direct and list experiment abortion prevalence estimates using subsets of data based on response times to list experiment related questions






List experiment control and treatment question response time
20
15
10
5
0
Abortion prevalence
10th
25th
50th
75th
90th
95th
Response time percentile
Direct estimate
List estimate

6. Aim 3: The Role of Familiarity in Reporting of Abortion on Surveys

6.1 Background

Social desirability pressures often cause respondents to significantly underreport sensitive or stigmatized behaviors on surveys (Tourangeau and Yan 2007). Induced abortion is considered a sensitive topic in nearly all settings, leading to substantial underestimates of this reproductive behavior. For example, investigators determined that respondents reported fewer than 50% of known abortions on a nationally representative reproductive health survey in the United States (Jones and Kost 2007). As such, researchers have sought to identify aspects of survey design that could improve reporting of abortion.

Performance Monitoring and Accountability 2020 (PMA2020) presents two largely untested survey design elements that could increase reporting of sensitive behaviors. The first design decision is the use of resident enumerators (REs) to interview respondents. The second is the repeated surveying of communities in 6 to 12 month intervals, resulting in a subset of the population being randomly selected for participation more than once. Both design elements result in greater interviewer and respondent familiarity, with the latter also contributing to a broader sense of familiarity with surveys and response confidentiality. We hypothesize both survey design features may increase respondents' likelihood of disclosing a prior abortion.

Researchers have long adhered to the idea that interviewers should be unknown to the respondent (Weinreb 2006). The stranger-interviewer norm relies on the assumption that familiarity between the interviewer and the respondent would negatively impact the validity of survey responses. Despite limited empirical evidence to support this idea, the stranger-interviewer model has been a mainstay of demographic surveys in low- and middle-income countries (LMICs). The Demographic and Health Survey (DHS) is the largest survey effort of its kind, and it ascribes to this paradigm by employing interviewers from outside the communities where they work. Part of the impetus for this design is practical, as finding qualified potential interviewers in many parts of LMICs continues to be difficult, if not impossible. But this is decreasingly the case as has been demonstrated by PMA2020's reliance on a cadre of REs to implement the project (Zimmerman et al. 2017). In most PMA2020 countries, the vast majority of interviewers live in or very near the enumeration area (Hawes et al. 2017).

The limited existing research on this topic demonstrates that prior respondent acquaintance with the interviewer does not reduce data quality and may actually improve it. Several studies out of the Dominican Republic found that interviewer-respondent familiarity had no effect on response rates or responses for nearly all variables investigated, and familiarity actually reduced non-response and significantly improved response validity on others (Rodriguez et al. 2015; Sana et al. 2016). Specifically, Rodriguez *et al.* found that interviewer-respondent familiarity had no effect on the reporting of sensitive information, including whether the respondent ever had an induced abortion, when reported in the context of a self-administered portion of the questionnaire (Rodriguez et al. 2015). Results from Sana *et al.* provided a consistent rejection of the stranger-interviewer norm, finding no effect of interviewer-respondent familiarity in 16 of the 18 questions investigated (Sana et al. 2016). For the two questions in which they detected a significant effect, respondents were *less* truthful when interviewed by an outsider (Sana et al. 2016). Using the same data, other investigators sought to determine whether there was a distinction between interviewers who were: 1) from the community and knew the respondent, 2) from the community and did not know the respondent, and 3) not from the community and did not know the respondent (Stecklov et al. 2015). Results indicated that local interviewers, regardless of familiarity with the specific respondent, obtained more realistic data on female sterilization than interviewers from outside the community (Stecklov et al. 2015).

More recently, researchers at PMA2020 investigated the potential impact of using REs and repeated cross-sectional face-to-face surveys to monitor family planning indicators in several sub-Saharan Africa countries (Safi et al. 2017). Outcomes of interest for this study included reporting of current modern contraceptive use, whether one has ever given birth, whether one has recently heard about family planning in the media, and whether one was visited by a health worker who discussed family planning in the past year (Safi et al. 2017). Findings demonstrated that RE-respondent acquaintance and participation in a prior round's survey do not appear to detrimentally impact data quality and in some contexts may actually improve it

(Safi et al. 2017). Specifically with regard to RE-respondent familiarity, being acquainted was associated with 1.76 (95% CI 1.41-2.19) and 1.15 (95% CI 1.03-1.29) times the odds of reporting current modern contraceptive use in Burkina Faso and Kenya, respectively; similar analyses resulted in null findings in Ghana, Ethiopia, and Uganda (Safi et al. 2017). Participation in a prior PMA2020 survey round was associated with 1.12 (95% CI 1.00-1.25) times the odds of reported modern contraceptive use in Kenya; findings were null in all other countries (Safi et al. 2017). One limitation of these analyses is the fact that social desirability for modern contraceptive use and other outcomes may not uniformly result in *under*reporting. This non-monotonic social desirability bias may partially explain the heterogeneity of estimates across country contexts.

The current chapter seeks to advance existing research on the stranger-interviewer norm with regard to abortion reporting. We aim to determine the potential impact of interviewer-respondent acquaintance and participation in a prior survey round on induced abortion reporting in Rajasthan, India. We hypothesize that both prior acquaintance with the interviewer and prior participation in a family planning survey will be independently associated with greater willingness to disclose an induced abortion, accounting for respondent, interviewer, and community characteristics. Results will provide evidence as to the role of these survey design elements in potentially improving reporting of induced abortion in India and other similar settings.

6.2 Methodology

6.2.1 Data

To investigate these relationships, we used PMA2020 data from Rajasthan, India (Zimmerman et al. 2017). The Indian Institute of Health Management and Research (IIHMR) collected the data in April and May of 2017, with technical guidance provided by the Bill & Melinda Gates Institute for Reproductive Health at the Johns Hopkins Bloomberg School of Public Health. We used a probabilistic two-stage cluster sampling design with probability proportional to size selection of enumeration areas (EA) from within urban/rural and region sampling domains. Thirty-five households were randomly selected from each EA and asked to participate in a household survey. REs invited all reproductive age women identified in a given household to participate in a female specific survey. Separately, the supervisors and REs conducted surveys at health service delivery points (SDP) that served the selected EAs. GPS points were taken with each household, female, and SDP survey. Interviewers asked for consent from all participants prior to administering surveys. The Institutional Review Boards (IRBs) at the Johns Hopkins Bloomberg School of Public Health and IIHMR provided ethical approval. Additional details regarding the design, sampling strategy, piloting, and training can be found in Chapter 3.

The outcome for this investigation was respondent report of abortion based on the direct abortion questions. The independent variables of interest were dichotomous measures of RE-respondent acquaintance and respondent participation in the prior PMA2020 round. RE-respondent acquaintance is information provided by the RE at the outset of the interview. The RE chooses from response options "very well acquainted", "well acquainted", "not well acquainted", and "not at all acquainted", which we then dichotomized by combining the two most-acquainted response options and the two least-acquainted response options. For the other independent variable, REs asked respondents whether they participated in the prior PMA2020 survey round.

Covariates of interest included respondent, RE, and community characteristics. Respondent characteristics of interest included age, marital status, education, wealth quintile, caste, religion, residence, and parity, all of which we treated as categorical variables for univariate and bivariate analyses and category specific indicator variables for multivariate analyses. The RE characteristics of interest included RE age, marital status, education, parity, their survey experience prior to PMA2020, whether they participated in the prior PMA2020 round, and whether they thought abortion was legal under any circumstances. We similarly treated these as categorical or indicator variables in analyses. We generated the community characteristics variables using multiple approaches. For EA modern contraceptive prevalence (MCP), EA average parity, and EA abortion prevalence we calculated the value for each respondent as cluster means excluding the respondent. Additionally, EA MCP and EA abortion prevalence were scaled to represent a 10% increase in the corresponding cluster level prevalence. For EA average distance to the nearest post-abortion care (PAC) providing facility, abortion providing facility, and pharmacy, we first used the SDP and female surveys' GPS data to calculate the Euclidian distance from each respondent to the nearest of each type of SDP and then created corresponding dichotomous variables to indicate whether the nearest of each type of SDP was less than 5 kilometers (km) away. We then took the average of the dichotomous variables for each EA.

6.2.2 Analysis

We initially conducted univariate analyses, exploring distributions of respondent, RE, and community characteristics. We then conducted bivariate analyses investigating the relationship between the two independent variables, RE-respondent acquaintance and respondent participation in the prior PMA2020 survey round, and each of the respondent, RE, and community characteristics using design-based F-tests to assess statistical significance for categorical variables and t-tests to assess statistical significance for continuous variables.

In order to estimate independent effects of respondent-specific variables, we calculated cluster means for each variable and then subtracted the cluster mean from the individual respondent answer to generate cluster mean centered variables. The cluster mean centered variables isolate the within interviewer effect of a given respondent level variable while the mean centered variables represent the between interviewer effect of cluster composition of respondent characteristics (Begg and Parides 2003). We used cluster mean centered versions of the independent variables and respondent characteristics in our multivariate analyses as we were interested in the within interviewer effect of these variables.

For multivariate analyses we used multilevel models with RE (interviewer) random effects. To determine the relative contribution of the sets of variables to explaining the variability in the observed data, we systematically added variables in the following manner: Model 1) only random effects; Model 2) added independent variables; Model 3) added respondent characteristics; Model 4) added RE characteristics; Model 5) added community characteristics (excluding EA abortion prevalence); and Model 6) added EA abortion prevalence. Thus, each model built onto the model prior. The six models took the following form:

1: 5: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\varsigma_j$

2: 5: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\beta_1\left(x_{1ij}-\bar{x}_{1i.}\right)+\beta_2\left(x_{2ij}-\bar{x}_{2i.}\right)+\varsigma_j$

3: 5: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\beta_1\left(x_{1ij}-\bar{x}_{1i.}\right)+\beta_2\left(x_{2ij}-\bar{x}_{2i.}\right)+\boldsymbol{\beta_3}\left(\boldsymbol{X_{ij}}-\boldsymbol{\bar{X}_{i.}}\right)+\varsigma_j$

4: 5: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\beta_1\left(x_{1ij}-\bar{x}_{1i.}\right)+\beta_2\left(x_{2ij}-\bar{x}_{2i.}\right)+\boldsymbol{\beta_3}\left(\boldsymbol{X_{ij}}-\boldsymbol{\bar{X}_{i.}}\right)+$

$\boldsymbol{\beta_4 V_j}+\varsigma_j$

5: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\beta_1\left(x_{1ij}-\bar{x}_{1i.}\right)+\beta_2\left(x_{2ij}-\bar{x}_{2i.}\right)+\boldsymbol{\beta_3}\left(\boldsymbol{X_{ij}}-\boldsymbol{\bar{X}_{i.}}\right)+$

$\boldsymbol{\beta_4 V_j}+\boldsymbol{\beta_5 \overline{W}_{i.}}+\boldsymbol{\beta_6 \bar{Z}_{i.}}+\varsigma_j$

6: $logit\left(\Pr\{y_{ij}=1\}\right)=\beta_0+\beta_1\left(x_{1ij}-\bar{x}_{1i.}\right)+\beta_2\left(x_{2ij}-\bar{x}_{2i.}\right)+\boldsymbol{\beta_3}\left(\boldsymbol{X_{ij}}-\boldsymbol{\bar{X}_{i.}}\right)+$

$\boldsymbol{\beta_4 V_j}+\boldsymbol{\beta_5 \overline{W}_{i.}}+\boldsymbol{\beta_6 \bar{Z}_{i.}}+\boldsymbol{\beta_7 \bar{A}_{i.}}+\varsigma_j$

where $logit\left(\Pr\{y_{ij}=1\}\right)$ is the log of the probably of reporting an abortion on the direct questions for respondent *i* interviewed by RE *j*, β_0 is the population abortion proportion when all other variables are the reference category, x_{1ij} is the level-1 RE-respondent acquaintance exposure variable, $\bar{x}_{1i.}$ is the cluster mean of the level-1 RE-

respondent acquaintance variable, $x_{1ij} - \bar{x}_{1i.}$ is the cluster mean centered individual RE-respondent acquaintance, β_1 is the within interviewer effect of RE-respondent acquaintance, x_{2ij} is the level-1 respondent participation in the prior PMA2020 survey exposure variable, $\bar{x}_{2i.}$ is the cluster mean of the level-1 respondent participation in the prior PMA2020 survey variable, $x_{2ij} - \bar{x}_{2i.}$ is the cluster mean centered individual respondent participation in the prior PMA2020 survey, β_2 is the within interviewer effect of respondent participation in the prior PMA2020 survey, X_{ij} is the vector of level-1 respondent covariates, $\bar{X}_{i.}$ is the vector of cluster mean level-1 respondent covariates, $X_{ij} - \bar{X}_{i.}$ is the vector of cluster mean centered level-1 respondent characteristics, β_3 is the vector of coefficients representing the within interviewer effect of a respondent characteristic, V_j is the level-2 level vector of RE covariates, β_4 is the corresponding vector of coefficients representing the between effect of a given RE characteristic, $\bar{W}_{i.}$ is the vector of cluster mean SDP-respondent dichotomous distance variables, β_5 is the vector of coefficients representing the between EA effect of a given SDP-respondent accessibility, $\bar{Z}_{i.}$ is the vector of level-1 characteristics averaged to represent community characteristics, β_6 is the vector of coefficients representing the between effect of a community's composition, $\bar{A}_{i.}$ is the level-1 abortion direct report averaged to represent the EA abortion prevalence, β_7 represents the between effect of EA abortion prevalence, and ς_j is the RE specific random effect. The random effects model assumptions took the following form: $E(\varepsilon_{ij}|\varsigma_j = 0)$, $E(\varsigma_j) = 0$, $Var(\varepsilon_{ij}|\varsigma_j) = \theta$, $Var(\varsigma_j) = \psi$ where ς_j is a random intercept for a given RE and they are assumed to be uncorrelated across REs and uncorrelated with level-1 residuals.

For all univariate and bivariate analyses, we incorporated survey weights and accounted for the sampling design using the Taylor linearization method. We used robust standard errors to account for clustering in the multilevel models after confirming standard error heteroskedasticity using the Breusch-Pagan test. We conducted all analyses in Stata version 15 (StataCorp 2017) and determined statistical significance using an alpha of 0.05.

We conducted a number of sensitivity analyses to test the robustness of our findings. We tested the impact of coding our independent variables differently, generating variables on whether the RE and/or respondent participated in the previous round, and generating variables regarding RE-respondent acquaintance and RE residence in relation to the EA (i.e. in, near, or far). We also tested different model specifications, including the number of integration points and the use of sampling weights.

6.3 Results

A total of 6,041 women completed the female questionnaire. On average, each RE completed female surveys with 41 respondents from their assigned EA. Among the 147 REs, only 118 completed the post data collection survey thus we present the RE characteristics for these 118 REs in Table 6.1. Mean RE age was 25.5 years, 68.6% of whom were currently married or cohabiting at the time of the survey. Forty-one percent of REs had primary, secondary, or technical training whereas 33.1% had graduated from or begun attending university and 26.3% had graduated from or begun a masters or doctoral program. More than half (53.4%) of REs had 1 to 2 children and 37.3% were nulliparous. Nearly one-third (32.2%) had prior survey experience and 81.4% had participated in PMA2020 Round 1 in Rajasthan. Most (58.1%) REs resided inside the EA in which they worked while 21.4% lived less than 10 km away and 19.7% lived further. Specifically relevant to the current investigation, 33% of REs thought abortion was not legal in India under any circumstances.

In Table 6.2a through 6.2c we present the percent distribution of the sample by variable groups and RE-respondent acquaintance; in Tables 6.3a through 6.3c we present the percent distribution of the sample by the same variable groups but by respondent participation in the prior PMA2020 survey round. Tables 6.2a and 6.3a contain the sample distribution by respondent characteristics and the independent variables. Overall, REs report being acquainted with 61.1% of respondents and 12.7% of respondents participated in the prior PMA2020 survey round (data not shown). Only 3.5% of respondents reported a prior induced abortion on the direct abortion questions. Fifty-five percent of women were below the age of 30, 75.7% were currently married or cohabiting, and many (36.7%) had never attend school. Nearly four in ten women (39.3%) were from an other backward caste and 85.3% resided in a Hindu household. Rajasthani women resided primarily in rural areas (64.1%) and more than one-third (36.1%) had 1 to 2 children. No respondent characteristics were statistically significantly associated with acquaintance or prior PMA2020 survey participation when using design-based F-test p-values but marital status, school, wealth, and residence were associated with RE-respondent acquaintance when conducting logistic regression accounting for clustering at the interviewer level. All associations between respondent characteristics and participation in the prior PMA2020 survey round remained null when accounting for clustering.

We present the sample distribution by RE characteristics and RE-respondent acquaintance in Table 6.2b and respondent participation in the prior PMA2020 survey round in Table 6.3b. Overall, nearly three-quarters of respondents were interviewed by REs age 20 to 29 (72.9%) who were currently married or cohabiting (72.3%). REs who had survey experience prior to PMA2020 conducted 29.9% of total surveys but REs who participated in the prior PMA2020 survey round conducted 77.6% of total surveys. REs who resided within the bounds of their assigned EA conducted more than half (51.6%) of the interviews and one-third (33.8%) were conducted by an RE who thought abortion was not legal in India under any circumstances. Only RE residence with regard to assigned EA was significantly associated with acquaintance when using a design-based F-test, with 61.1% of surveys conducted by an RE who knew the respondent living within the EA compared to 36.0% of surveys conducted by an RE who did not know the respondent residing within the EA (p-value=0.01). RE age and marital status were also significantly associated with RE-respondent acquaintance when accounting for clustering, with younger women and never married women being more likely to be acquainted with their interviewee. No RE characteristics were associated with respondent participation in the prior PMA2020 survey round regardless of statistical approach.

Tables 6.2c and 6.3c contain the sample distribution by community characteristics and the independent variables. We see that on average women live in EAs where the modern contraceptive prevalence is 43.0%, but this varied significantly by acquaintance when using a design-based F-test; 39.4% among respondents not acquainted and 45.3% among those acquainted. This difference was not significant when accounting for clustering. Average parity across EAs was 1.9 while average abortion prevalence was 3.5%. With regard to distance to the nearest PAC providing facility, 24.4% of respondents lived in an EA where residents lived on average less than 5 km away. This proportion for distance to the nearest abortion providing facility and pharmacy was 19.3% and 52.0%, respectively. Only EA MCP differed significantly by acquaintance when assessing via design-based F-test, with 45.2% of women acquainted with their interviewer reporting current use of a modern contraceptive method as opposed to 39.4% of women who were not acquainted with their interviewer (p-value<0.001). EA MCP was not statistically significantly associated with acquaintance when accounting for clustering. With regard to respondent participation in the prior survey round, EA MCP was statistically significant when accounting for interviewer clustering, with those who participated having reported slightly greater use (45.5%) than those who did not (42.6%, p-value=0.04).

We present results from the multivariate models in Table 6.4. In the first model, we see that 42% of the variability in abortion reporting was explained by the clustering at the interviewer level, which is equivalent to clustering at the EA level. Adding the independent variables in Model 2, we see that neither RE-respondent acquaintance nor respondent participation in the prior PMA2020 survey round was associated with abortion reporting. Correspondingly, rho is unaffected; 42% of variability in the abortion reporting remains explained by clustering at the interviewer level. In Model 3 we include the cluster mean centered level-1 respondent characteristics. Women who were never married were significantly less likely to report a prior induced abortion (OR=0.09, 95% CI 0.02-0.49), while urban women (OR=3.48, 95% CI 1.69-7.15) and women with children (1-2 children OR=5.16, 95% CI 1.75-15.27; 3-4 children OR=6.95, 95% CI 2.37-20.36; 5+ children OR=6.33, 95% CI 1.93-20.76) were significantly more likely to report an induced abortion.

When we added RE characteristics in Model 4 we saw the variability in the dependent variable that is explained by the interviewer cluster reduced to 35%. While the independent variables remain insignificant in Model 4, never married, residence, and parity remain statistically significant and we see that being from an other backward caste was associated with significantly lower odds of abortion reporting (OR=0.52, 95% CI 0.28-0.99). Additionally, women who were interviewed by an RE with at least some university education were significantly more likely to report an induced abortion compared to those interviewed by an RE with secondary school or less education (OR=2.54, 95% CI 1.00-6.42).

In Model 5, we added all of the community characteristics except EA abortion prevalence. Only EA average distance to the nearest abortion providing facility was significantly associated with abortion reporting, with respondents living in an EA where all women were less than 5 km from an abortion providing facility having 6.35 times the odds of reporting an induced abortion (95% CI 1.23-32.80). Adding these variables also reduced remaining variability in abortion reporting explained by the interviewer cluster to 32%. In the final model, we add the EA abortion prevalence, which explains the remaining variability in abortion reporting attributable to the cluster (rho=0.00).

In Model 6 we see that women residing in an EA where the abortion prevalence is 10% higher had 3.44 times the odds of reporting and induced abortion (95% CI 2.48-4.76). EA average distance to an abortion providing facility remained statistically significant in this final model (OR=3.95, 95% CI 1.13-13.74), as did never being married, being from an other backward caste, living in an urban area, and having children (Table 6.4). RE-respondent acquaintance remained not statistically significant, although the odds of reporting an abortion were elevated for those who were acquainted with their interviewer (OR=1.79, 95% CI 0.60-5.34). Respondent participation in the previous PMA2020 survey round was similarly not significant in the final model (OR=1.11, 95% CI 0.67-1.83). The corrected AICs gradually decreased from 1,415 in Model 1 to 1,028 in Model 6 as we added explanatory variables, providing support to the fact that Model 6 best explains the observed data.

To test the robustness of our findings, we conducted a number of sensitivity analyses. Given the four acquaintance response options, we could have dichotomized the variable differently. Using only "not acquainted" as the reference category and coding any level of acquaintance ultimately yielded the same results, but the OR was closer to the null (OR=1.09, 95% CI 0.42-2.84). Investigating the total effect of acquaintance by using the non-cluster mean centered acquaintance variable, our results were again closer to the null than our primary findings (OR=1.32, 95% CI 0.70-2.49). Generating a six-category variable that combined whether the RE reported being acquainted with the respondent and whether the RE resided within, near, or far from the assigned EA also resulted in null findings. In the final models presented in Table 6.4, we accounted for the sampling design by adjusting for urban/rural strata and interviewer (or EA) clustering. Many researchers also include survey weights in analyses, although the necessity of this is debated. When including the female sampling weights as a continuous explanatory variable, our results were the same (RE-respondent acquaintance OR=1.79, 95% CI 0.62-5.21; respondent participation in prior PMA2020 survey round OR=1.12, 95% CI 0.68-1.83) and the sampling weight variable was not significant (OR=1.33, 95% CI 0.80-2.22). Additionally, we tested using different numbers of integration points in our multivariate models between 8 and 40 and confirmed that our results were stable. In conjunction, the results from these sensitivity analyses demonstrate our primary findings are robust to many model specifications.

6.4 Discussion

Findings provide further evidence to reject the stranger-interviewer norm in survey research. Specifically, we demonstrate that interviewer-respondent familiarity was not associated with reporting of abortion on a face-to-face survey in Rajasthan, India, nor was respondent participation in the prior survey round. Thus, we reject the null hypotheses that respondent familiarity with the interviewer or the survey process would be associated with greater comfort in revealing sensitive behaviors, like abortion. These findings are largely consistent with existing literature that found familiarity between respondents and interviewers did not reduce data quality, and instead, may have increased it in some instances (Rodriguez et al. 2015; Safi et al. 2017; Sana et al. 2016; Stecklov et al. 2015; Weinreb 2006). While our findings are specific to a survey implemented in Rajasthan, India, we believe this evidence, in conjunction with prior research, suggests this may be true for surveys implemented in other similar settings.

Unfortunately, our null findings indicate that researchers cannot leverage these survey design features to improve data on induced abortion. Thus, investigators must continue seeking to identify survey design decisions or methodologies to reduce social desirability pressure in abortion reporting.

Our secondary findings regarding characteristics associated with greater abortion reporting provide interesting new insights into the influence of women's community on their use of or willingness to report induced abortion. Generating cluster specific aggregate measures of individual level characteristics allowed us to quantify the potential independent association between the composition of a woman's community and abortion reporting, which to our knowledge had never been investigated. Specifically, an increase of 10% in the prevalence of abortion reported in a woman's community was associated with 3.44 (95% CI 2.48-4.76) times the odds of an individual woman from that community reporting an induced abortion. These results should be interpreted with caution, though, as this variable is essentially an aggregated form of the dependent variable.

Our ability to account for the service delivery environment characteristics also strengthened our findings and provided novel results regarding the influence that access plays on likelihood of abortion and associated abortion reporting. We found that residing in a community where women on average live less than 5 km from a facility that provides abortion is associated with 3.95 (95% CI 1.13-13.74) times the odds of reporting a prior induced abortion. Further analyses are required to better understand the spatial relationship between induced abortion and the service delivery environment, as well as the potential inequities in safe abortion access these results reveal.

This study has a number of strengths. The data we collected include survey responses from more than 6,000 women. This large sample size and diversity of responses enabled us to test complex models. The representative data included questions on a range of socioeconomic and reproductive health topics, allowing us to adjust for many potential confounding variables. Additionally, because PMA2020 contemporaneously collects SDP data, we were able to investigate the role of women's access to abortion or PAC services, as well as pharmacies, where women commonly access medical abortion pills (Singh et al. 2017). Unlike reporting of other potentially sensitive behaviors, for instance current use of modern contraception, induced abortion has a reliably monotonic bias. This means our null findings are not the result of some women perhaps being *more* likely to report an abortion as a result of social desirability pressure; the stigma surrounding abortion consistently works to reduce reporting of induced abortion. We also conducted several sensitivity analyses that illustrated our findings are robust.

However, this study also has a number of limitations. The primary weakness is the conflation of abortion experience and abortion reporting. The dependent variable is a combination of actual prior abortion experience and willingness to report it in a survey. We tried to isolate the effect of RE-respondent reporting and prior participation in the PMA2020 survey round by adjusting for several respondent and community characteristics that are associated with experience of abortion. We feel this allowed us to isolate the independent effect of our independent variables on abortion *reporting*. Related to our independent variables, these might not be capturing the most important aspects of familiarity. Our acquaintance variable may not be measuring the RE-respondent familiarity appropriately, and other features of familiarity, like the type of relationship (e.g. familial, similarly aged friend, etc.) may be more critical to the reporting of sensitive behaviors. And respondent participation in any prior survey or just being familiar with the PMA2020 project, regardless of participating in the prior PMA2020 survey round, may have more influence on abortion reporting. However, given the robustness of our findings to different measures of acquaintance, we feel this limitation is minor and feel confident our results would not change qualitatively. Additionally, our use of Euclidean distance in generating the service delivery environment variables may not be the appropriate approach. Further investigation is required.

Our findings provide support to the use of survey models like PMA2020, which trains women from the sampled communities to conduct repeated cross-sectional surveys in the areas where they live (Zimmerman et al. 2017); data do not appear to be affected by these design features. Thus, if using resident enumerators can reduce costs or the challenges associated with entry into a local community, researchers could implement the PMA2020 approach more broadly. However, continued investigation of the aspects of familiarity we analyzed here is needed to confirm whether these findings can be generalized to other social settings.

Unfortunately, the quest continues to identify aspects of survey design that could improve the validity of abortion reporting. While we had hypothesized that respondent familiarity with the interviewer or the survey experience would increase their reporting of sensitive behaviors, namely abortion, we did not observe this phenomenon. In future rounds of data collection, we intend to test different question methodologies in continuance of our efforts to improve abortion reporting on surveys. Recent supply side and indirect measurement of abortion in India confirmed that abortion is being widely used by women to control their fertility (Singh et al. 2017). Investigators estimate that there are 47 abortions per 1,000 women age 15 to 49 annually, 73% of which are medical abortions done outside of health facilities (Singh et al. 2017). These estimates, calculated using largely facility and drug distribution data, do not provide information on the social epidemiology of abortion or who is most at risk for an unintended pregnancy that ends in abortion. Researchers must steadfastly seek to understand the demographics of women most at risk for unsafe abortion related morbidity and mortality and to provide programmatic and policy relevant data in order to improve the delivery of family planning and abortion services.

Table 6.1. Characteristics of Rajasthani REs[1]		
	%	N
Mean age (SE)	25.5 (0.6)	118
Marital status		
Currently married/cohabiting	68.6	81
Divorced or separated/widowed	2.5	3
Never married	28.8	34
School		
Primary, secondary, or technical	40.7	48
University	33.1	39
Masters or doctoral	26.3	31
Parity		
0	37.3	44
1-2	53.4	63
3-4	9.3	11
5+	0.0	0
Prior survey experience		
No	67.0	79
Yes	32.2	38
Residence with regard to assigned EA		
In	58.1	68
Near (<=10 km)	21.4	25
Far (> 10 km)	19.7	23
Participated in previous PMA2020 survey		
No	18.6	22
Yes	81.4	96
Thinks abortion not legal under any circumstances		
No	67.0	77
Yes	33.0	38
Total	100.0	118

[1]Only 118 of the 147 REs completed the RE characteristics survey following completion of data collection activities

Table 6.2a. Percent distribution of sample by respondent characteristics and respondent/RE acquaintance[1]					
Respondent characteristics	Not Acquainted	Acquainted	Total	p-value[2]	p-value[3]
Abortion (direct question)					
No	96.0	96.8	96.5	0.58	0.23
Yes	4.0	3.2	3.5		
Age					
15-19	19.0	18.9	18.9	0.66	0.19
20-29	37.1	35.5	36.1		
30-39	26.3	27.5	27.0		
40-49	17.6	18.2	18.0		
Marital status					
Currently married/cohabiting	77.1	74.8	75.7	0.20	**0.03**
Divorced or separated/widowed	2.8	2.6	2.7		
Never married	20.1	22.6	21.6		
School					
Never attended	38.6	35.5	36.7	0.07	**0.03**
Primary	24.4	24.4	24.4		
Secondary	19.1	16.6	17.6		
Higher or postgraduate	17.9	23.5	21.3		
Wealth					
Poorest	16.7	16.1	16.3	0.67	**<0.01**
Second poorest	17.7	17.4	17.5		
Middle	19.5	19.8	19.7		
Second wealthiest	23.5	20.3	21.5		
Wealthiest	22.7	26.4	24.9		
Caste of household head					
Scheduled caste	20.8	23.4	22.4	0.63	0.51
Scheduled tribe	18.9	15.9	17.0		
Other backward caste	36.8	40.9	39.3		
General	23.6	19.8	21.3		
Religion of household head					
Hindu	85.7	84.9	85.3	0.49	0.28
Muslim	13.6	13.1	13.3		
Other	0.6	1.9	1.4		
Residence					
Rural	58.7	67.5	64.1	0.27	**<0.01**
Urban	41.3	32.5	35.9		
Parity					
0	30.5	31.4	31.0	0.84	0.44
1-2	36.4	35.9	36.1		
3-4	24.5	24.8	24.7		
5+	8.6	7.9	8.2		

[1]All estimates include weights accounting for complex survey design and non-response

[2]P-value from design-based f-test for categorical variables and t-test for continuous variables
[3]P-value from logistic regression and associated chi-squared test

Table 6.2b. Percent distribution of sample by RE characteristics and respondent/RE acquaintance[1]

RE characteristics	Not Acquainted	Acquainted	Total	p-value[2]	p-value[3]
Age					
15-19	7.5	10.0	9.0	0.20	**0.05**
20-29	66.5	76.9	72.9		
30-39	20.5	10.6	14.4		
40+	5.5	2.5	3.7		
Marital status					
Currently married/cohabiting	77.6	69.0	72.3	0.07	**<0.01**
Divorced or separated/widowed	3.0	0.6	1.5		
Never married	19.4	30.4	26.2		
School					
Primary, secondary, or technical	36.1	42.8	40.3	0.32	0.44
University	39.6	27.6	32.2		
Masters or doctoral	24.3	29.6	27.6		
Parity					
0	28.1	43.9	37.9	0.11	0.16
1-2	58.0	48.6	52.2		
3-4	13.9	7.5	9.9		
Prior survey experience					
No	65.7	72.8	70.1	0.33	0.29
Yes	34.3	27.2	29.9		
Residence with regard to assigned EA					
In	36.0	61.1	51.6	**<0.01**	**<0.01**
Near (<=10 km)	29.4	22.6	25.2		
Far (> 10 km)	34.6	16.4	23.3		
Participated in previous PMA2020 survey					
No	28.2	19.2	22.4	0.25	0.08
Yes	71.8	80.8	77.6		
Thinks abortion not legal under any circumstances					
No	64.1	67.5	66.2	0.66	0.83
Yes	35.9	32.5	33.8		

[1]All estimates include weights accounting for complex survey design and non-response
[2]P-value from design-based f-test for categorical variables and t-test for continuous variables
[3]P-value from logistic regression and associated chi-squared test

Table 6.2c. Percent distribution of sample by community characteristics and respondent/RE acquaintance[1]

Community characteristics	Not Acquainted	Acquainted	Total	p-value[2]	p-value[3]
EA average distance to nearest PAC providing facility <5 km	22.1	27.6	24.4	0.47	0.57
EA average distance to nearest abortion providing facility <5 km	21.6	17.9	19.3	0.59	0.34
EA average distance to nearest pharmacy <5 km	55.8	49.6	52.0	0.43	0.20
EA modern contraceptive prevalence	39.4	45.2	43.0	**<0.01**	0.17
EA average parity	1.9	2.0	1.9	0.52	0.79
EA abortion prevalence	4.1	3.0	3.5	0.42	0.14

[1]All estimates include weights accounting for complex survey design and non-response

[2]P-value from design-based f-test for categorical variables and t-test for continuous variables

[3]P-value from logistic regression and associated chi-squared test

Table 6.3a. Percent distribution of sample by respondent characteristics and respondent participation in the previous PMA2020 survey[1]					
Respondent characteristics	No Prior Survey	Prior Survey	Total	p-value[2]	p-value[3]
Abortion (direct question)					
No	96.5	96.1	96.5	0.65	0.77
Yes	3.5	3.9	3.5		
Age					
15-19	19.1	17.7	18.9	0.15	0.12
20-29	36.1	35.9	36.1		
30-39	27.3	25.1	27.0		
40-49	17.5	21.3	18.0		
Marital status					
Currently married/cohabiting	76.0	72.7	75.7	0.21	0.06
Divorced or separated/widowed	2.6	3.4	2.7		
Never married	21.4	23.9	21.6		
School					
Never attended	37.4	32.0	36.7	0.20	0.21
Primary	24.0	26.3	24.4		
Secondary	17.5	18.0	17.6		
Higher or postgraduate	21.1	23.7	21.3		
Wealth					
Poorest	17.0	12.6	16.3	0.36	0.64
Second poorest	16.9	20.8	17.5		
Middle	19.5	20.8	19.7		
Second wealthiest	21.6	21.0	21.5		
Wealthiest	25.0	24.9	24.9		
Caste of household head					
Scheduled caste	22.0	24.9	22.4	0.44	0.42
Scheduled tribe	16.7	19.7	17.0		
Other backward caste	39.4	38.4	39.3		
General	22.0	17.0	21.3		
Religion of household head					
Hindu	85.4	84.2	85.3	0.81	0.67
Muslim	13.2	14.6	13.3		
Other	1.5	1.2	1.4		
Residence					
Rural	64.0	63.8	64.1	0.98	0.66
Urban	36.0	36.2	35.9		
Parity					
0	31.0	31.3	31.0	0.85	0.87
1-2	36.0	37.0	36.1		
3-4	24.9	23.2	24.7		
5+	8.2	8.4	8.2		

[1]All estimates include weights accounting for complex survey design and non-response

[2]P-value from design-based f-test for categorical variables and t-test for continuous variables

[3]P-value from logistic regression and associated chi-squared test

Table 6.3b. Percent distribution of sample by RE characteristics and respondent participation in the previous PMA2020 survey[1]

RE characteristics	No Prior Survey	Prior Survey	Total	p-value[2]	p-value[3]
Age					
15-19	9.9	3.1	9.0	0.06	0.28
20-29	71.4	82.6	72.9		
30-39	15.3	9.0	14.4		
40+	3.4	5.2	3.7		
Marital status					
Currently married/cohabiting	71.6	76.1	72.3	0.37	0.32
Divorced or separated/widowed	1.5	1.3	1.5		
Never married	26.8	22.6	26.2		
School					
Primary, secondary, or technical	40.3	40.8	40.3	0.94	0.31
University	32.2	30.5	32.2		
Masters or doctoral	27.5	28.7	27.6		
Parity					
0	38.5	33.0	37.9	0.44	0.90
1-2	51.3	58.6	52.2		
3-4	10.2	8.4	9.9		
Prior survey experience					
No	70.5	67.5	70.1	0.58	0.47
Yes	29.5	32.5	29.9		
Residence with regard to assigned EA					
In	50.5	59.0	51.6	0.23	0.88
Near (<=10 km)	24.9	25.1	25.2		
Far (> 10 km)	24.5	16.0	23.3		
Participated in previous PMA2020 survey					
No	22.7	19.2	22.4	0.51	0.08
Yes	77.3	80.8	77.6		
Thinks abortion not legal under any circumstances					
No	65.1	73.2	66.2	0.12	0.43
Yes	34.9	26.8	33.8		

[1]All estimates include weights accounting for complex survey design and non-response

[2]P-value from design-based f-test for categorical variables and t-test for continuous variables

[3]P-value from logistic regression and associated chi-squared test

Table 6.3c. Percent distribution of sample by community characteristics and respondent participation in the previous PMA2020 survey[1]

Community characteristics	No Prior Survey	Prior Survey	Total	p-value[2]	p-value[3]
EA average distance to nearest PAC providing facility <5 km	25.0	28.9	24.4	0.45	0.57
EA average distance to nearest abortion providing facility <5 km	18.8	23.3	19.3	0.35	0.34
EA average distance to nearest pharmacy <5 km	51.9	52.5	52.0	0.91	0.50
EA modern contraceptive prevalence	42.6	45.5	43.0	0.06	**0.04**
EA average parity	2.0	1.9	1.9	0.50	0.73
EA abortion prevalence	3.5	3.4	3.5	0.96	0.98

[1]All estimates include weights accounting for complex survey design and non-response

[2]P-value from design-based f-test for categorical variables and t-test for continuous variables

[3]P-value from logistic regression and associated chi-squared test

Table 6.4. Odds of abortion reporting on respondent/RE acquaintance and respondent previous participation in a PMA survey and associated intracluster correlation adjusting for nothing (Model 1), independent variables (Model 2), respondent characteristics (Model 3), RE characteristics (Model 4), community characteristics (Model 5), and EA abortion prevalence (Model 6)

	Model 1			Model 2			Model 3			Model 4			Model 5			Model 6		
	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Independent variables																		
Respondent/RE acquainted (reference no)				1.66	0.73	3.77	1.78	0.76	4.18	1.78	0.62	5.13	1.74	0.59	5.13	1.79	0.60	5.34
Respondent participated in previous round				1.01	0.65	1.59	1.01	0.64	1.60	1.15	0.69	1.91	1.14	0.69	1.90	1.11	0.67	1.83
Respondent characteristics																		
Age (reference 15-19)																		
20-29							1.02	0.41	2.58	0.83	0.27	2.57	0.81	0.26	2.53	0.82	0.26	2.57
30-39							1.22	0.56	2.63	1.12	0.44	2.88	1.07	0.41	2.79	1.11	0.42	2.90
40-49							0.55	0.25	1.21	0.48	0.18	1.27	0.47	0.17	1.25	0.48	0.17	1.34
Marital status (reference currently																		
Divorced/widowed							0.58	0.18	1.85	0.80	0.27	2.42	0.81	0.27	2.46	0.83	0.29	2.40
Never married							**0.09**	**0.02**	**0.49**	**0.15**	**0.02**	**0.93**	0.14	0.02	1.03	**0.16**	**0.03**	**0.92**
School (reference never attended)																		
Primary							1.23	0.84	1.81	1.10	0.69	1.75	1.12	0.70	1.78	1.12	0.71	1.77
Secondary or higher							1.54	0.86	2.76	1.74	0.93	3.28	1.78	0.95	3.33	1.83	0.99	3.37
Wealth quintile (reference poorest)																		
Middle poorest							0.61	0.33	1.14	0.40	0.19	0.84	**0.38**	**0.18**	**0.80**	**0.39**	**0.20**	**0.77**
Middle							0.75	0.37	1.54	0.59	0.25	1.36	0.53	0.23	1.20	0.53	0.24	1.16
Middle wealthiest							0.59	0.27	1.32	0.53	0.21	1.32	0.49	0.20	1.20	0.49	0.20	1.18
Wealthiest							0.62	0.26	1.48	0.57	0.21	1.57	0.53	0.19	1.44	0.51	0.20	1.34
Caste (reference scheduled caste)																		
Scheduled tribe							0.57	0.25	1.28	0.56	0.19	1.67	0.58	0.19	1.72	0.60	0.21	1.77
Other backward caste							0.57	0.33	1.01	**0.52**	**0.28**	**0.99**	**0.50**	**0.27**	**0.94**	**0.51**	**0.28**	**0.94**
Generate caste							1.01	0.55	1.88	0.95	0.48	1.86	0.94	0.48	1.85	0.95	0.47	1.90
Religion (reference Hindu)																		
Muslim							1.49	0.55	4.05	1.05	0.29	3.79	1.05	0.28	3.88	1.06	0.28	3.96
Other religion							1.57	0.76	3.26	1.34	0.60	3.03	1.35	0.60	3.01	1.32	0.61	2.85
Residence (reference rural)																		
Urban							**3.48**	**1.69**	**7.15**	**5.88**	**2.65**	**13.07**	**5.20**	**1.90**	**14.21**	**2.24**	**1.29**	**3.89**
Parity (reference 0)																		
1-2							**5.16**	**1.75**	**15.27**	**7.48**	**2.77**	**20.15**	**7.98**	**2.79**	**22.81**	**7.35**	**2.68**	**20.14**
3-4							**6.95**	**2.37**	**20.36**	**10.36**	**3.87**	**27.71**	**11.62**	**4.16**	**32.47**	**10.45**	**3.95**	**27.68**
5+							**6.33**	**1.93**	**20.76**	**11.29**	**3.71**	**34.30**	**13.05**	**4.14**	**41.13**	**11.35**	**3.83**	**33.57**

Table 6.4. Continued

RE characteristics																		
Age (reference 15-19)																		
20-29										0.44	0.13	1.48	0.49	0.15	1.56	0.99	0.54	1.81
30-39										0.42	0.08	2.31	0.43	0.08	2.50	0.81	0.31	2.16
40+										0.74	0.06	8.97	0.86	0.11	6.88	1.23	0.49	3.09
Married ever (reference no)																		
Yes										0.60	0.18	2.00	0.68	0.20	2.25	0.92	0.45	1.88
School (primary, secondary, or technical)																		
University										**2.54**	**1.00**	**6.42**	**2.85**	**1.12**	**7.23**	1.62	0.89	2.96
Masters or doctoral										0.91	0.35	2.41	0.82	0.32	2.09	0.69	0.37	1.27
Parity (reference 0)																		
1-2										0.59	0.19	1.87	0.63	0.20	1.96	0.74	0.37	1.48
3-4										1.87	0.41	8.61	1.89	0.42	8.57	1.18	0.57	2.46
Residence with regard to assigned EA (reference in)																		
Near (<=10 kms)										0.75	0.27	2.08	0.80	0.31	2.10	0.88	0.48	1.63
Far (> 10 kms)										0.54	0.21	1.34	0.54	0.22	1.33	0.73	0.47	1.13
Participated in previous PMA2020 survey (reference no)																		
Yes										0.63	0.29	1.39	0.63	0.27	1.49	0.82	0.54	1.25
Thinks abortion not legal under any																		
Yes										0.78	0.35	1.74	0.77	0.35	1.70	0.85	0.52	1.38
Community characteristics																		
EA average distance to nearest PAC providing facility <5 km													0.27	0.06	1.16	0.34	0.09	1.21
EA average distance to nearest abortion providing facility <5 km													**6.35**	**1.23**	**32.80**	**3.95**	**1.13**	**13.74**
EA average distance to nearest pharmacy <5 km													1.28	0.61	2.65	1.00	0.65	1.54
EA modern contraceptive prevalence increase of 10%													1.04	0.82	1.33	1.06	0.91	1.23
EA average parity													1.57	0.57	4.31	1.07	0.60	1.94
EA reported abortion prevalence increase of 10%																**3.44**	**2.48**	**4.76**
Rho	0.42			0.42			0.42			0.35			0.32			0.00		
Corrected AIC	1,415			1,416			1,313			1,037			1,031			1,028		
N	6,041			6,016			5,993			4,984			4,951			4,951		

7. Discussion

7.1 Summary of Findings

Induced abortion is a ubiquitous yet elusive phenomenon. It is among the most common health experiences, yet despite investigators' best efforts, we know relatively little about the specifics of its occurrence in most low-resource settings. This is increasingly the case with widespread access to medical abortion drugs in most parts of the world. Women can now use these drugs to self-induce with relative ease and safety regardless of legality and access to services in the formal healthcare system. As such, abortion presents among the most difficult measurement and epidemiological challenges.

We sought to identify features of interviewer administered surveys that could reduce bias in abortion reporting. The broad aim of this work was to enable greater understanding of the extent and social epidemiology of abortion and unintended pregnancy in Rajasthan, India. In our investigation of abortion reporting and underreporting, we examined aspects of survey design and question methodology that could impact women's willingness to report abortion on face-to-face surveys.

In Aim 1, we tested the use of a list experiment to measure induced abortion prevalence (and incidence). While recent evidence suggested the list experiment may produce significantly more valid reporting of induced abortion (Moseson et al. 2015), our results did not produce better estimates than direct questions. Despite having a large sample size of reproductive age women, the list experiment estimate of lifetime experience of abortion was actually significantly *less* valid than the direct abortion estimate (1.8% versus 3.5%). Further investigation into the list experiment assumptions revealed evidence of significant design effects. As such, women responded differently to the control items on the list when they encountered these items with and without the induced abortion item. In sum, the confidentiality afforded by the list experiment questions did not overcome women's desire for privacy.

In Aim 2, we explored specific mechanisms to explain the list experiment's failure to produce more valid estimates of induced abortion. Using survey paradata, we examined whether response times can illuminate list experiment failures at different stages of the response process. We did not find evidence that poor numeracy or poor cognitive ability (as measured by schooling) explains the list experiment's poor performance. However, we did identify a significant editing effect whereby women who reported an abortion on the direct questions took significantly longer to respond to the list experiment treatment list compared to women who reported no abortion on the direct questions. This is consistent with social and cognitive psychology theory and empirical evidence on the impact of social desirability pressure in responding to sensitive survey questions (Holtgraves 2004; Holtgraves et al. 1997; Sudman et al. 1996).

For Aim 3 we focused more broadly on the context in which respondents answer abortion questions. Specifically, we investigated whether interviewer-respondent acquaintance or respondent prior experience with the reproductive health survey would impact a woman's willingness to report a prior induced abortion. In both instances, we found these aspects of familiarity were not statistically significantly associated with abortion reporting, adjusting for respondent, interviewer, and community characteristics. These findings provide support for the PMA2020 RE model and demonstrate that PMA2020 data on abortion are likely to be comparable to data from stranger-interviewer surveys. But our results leave us unsatisfied in terms of our interest in identifying survey design features that would increase the validity of abortion reporting on face-to-face surveys.

7.2 Overall Strengths

This dissertation research has a number of strengths. Results from this investigation provide a thorough investigation of abortion reporting and underreporting on surveys in low-resource settings. This is among the most in-depth studies of this phenomenon and the social and cognitive processes involved. Many of the specific analyses constitute the first investigations exploring these relationships with regard to abortion. As such, we contribute significantly to the literature on survey based abortion reporting.

With regard to the specifics of the research design and data, this dissertation had a number of advantages. The list experiment methodology we used is still novel in its application to abortion prevalence and incidence measurement. We added the list experiment questions to a population-based survey, which was the largest sample in which list experiment questions had ever been implemented, to our knowledge. This minimized concerns that poor performance could be explained by insufficient sample. The broader survey data provided a wealth of information from which we could explore relationships across demographic and reproductive characteristics. Using contemporaneously collected direct question data provided an opportunity for robust investigation of the continued bias present in the list experiment estimates. The use of paradata to investigate the social and cognitive processes involved in answering list experiment questions was innovate and provided new insights. Additionally, the PMA2020 model presented the opportunity to investigate the impact of survey design features that had no or only minimal prior testing.

7.3 Overall Limitations

While the proposed dissertation research has many strengths, there remains a number of limitations. The largest issue with any of the current means of estimating induced abortion in low-resource settings is the lack of an objective measure against which to validate results. We capitalized on the use of the direct questions estimates for comparison, in addition to a thorough assessment of design assumptions. Ideally, researchers would implement a true validation of the list experiment and other indirect methodologies using clinical data, which would provide strong evidence as to the validity of the methods. Comparison with a clinic-based sample could provide the necessary additional data regarding the type of women who are most likely to underreport in a given setting, which researchers could use to more accurately calibrate population-based survey results.

As with any survey data, there was a potential for bias as a result of incomplete listing frame coverage. If specific types of women were systematically missed or underrepresented by the PMA2020 sampling procedures, this could introduce bias in the resulting estimates. Random sampling error also could have resulted in incorrect and/or imprecise estimates by chance alone. In addition, due to the sensitive nature of abortion, underreporting clearly continued to be a concern in both the direct and list experiment abortion questions.

7.4 Implications

Our motivation to improve abortion measurement was manifold. Abortion is a common means of controlling one's fertility. We were interested in measuring abortion because it provides insight into a principal aspect of population dynamics. Accurately quantifying induced abortion tells us the extent to which women's contraceptive needs are being met and how much they are relying on abortion as a secondary means of fertility control. Measuring abortion is also important because it helps in accurately calculating rates of pregnancy, unintended pregnancy, and contraceptive failure (Finer and Zolna 2016; Sundaram et al. 2017).

In India, recent estimates from many states point to a decline in the total fertility rate and a concurrent decline in contraceptive use (International Institute for Population Sciences (IIPS) and Macro International 2015-2016). Abortion, and specifically unofficial use of medical abortion drugs, is likely playing a role in these trends. But the extent to which this is true remains unknown. Other factors that could explain these data include delays in the age of marriage, poor reporting of prior sterilization, or generally poor-quality data. Recent induced abortion estimates suggest that abortion rates are higher than previously thought, and that use of medical abortion drugs is widespread (Singh et al. 2017). Thus, initial evidence suggests abortion is likely contributing substantially to the seemingly contradictory fertility and contraceptive use trends, though further research is required. Better measurement of the demography and incidence of abortion, particularly unsafe abortion, can shed light on how this behavior fits in to women's repertoire of pregnancy management choices. Understanding which populations are most struggling to control their fertility through contraceptive use and safe abortion would allow government officials to focus limited resources to service improvement and related communications programs among these populations.

Even with good, accurate abortion data, interpreting results is not always straightforward. Research from the United States recently found that abortion rates have dramatically decreased in recent years (Jones and Jerman 2017). Results suggest the abortion rate declined 25%, from 19.4 abortions per 1,000 women age 15 to 44 in 2008 to 14.6 in 2014 (Jones and Jerman 2017). The abortion rate change for adolescents age 15 to 19 was the most marked, declining by 46% (Jones and Jerman 2017). Unfortunately, drawing conclusions from these trends is difficult (Foster 2017). A seemingly simple question regarding whether this finding demonstrates a "public health achievement or failure" remains unanswered (Foster 2017). Are women now more effectively contracepting and preventing unwanted pregnancies, or have the barrage of targeted regulations of abortion provider (TRAP) laws impeded women's access to abortion services? Are the declines real or is the extent of self-induction substantial enough to offset the observed declines in facility-based abortion (Grossman et al. 2014)? Further research is required to answer these and other pressing public health questions.

With regard to specific policy implications in India, government officials can better design programs seeking to address inequities in access to quality contraceptive and safe abortion services using the information produced by studies like those by Singh *et al*. Their results demonstrated women's overwhelming reliance on medical abortion drugs procured illegally from pharmacies to terminate (Singh et al. 2017). Increasing legitimate access to medical abortion by allowing a broader range of providers to administer medical abortion drugs could improve equity in access to reproductive health services.

There are steps the government can take to improve abortion data collection prospectively. The official government data are compiled from reported service statistics and extracted from logbooks in public and private facilities that are registered to provide abortion services. These registered facilities comprise only a small fraction of the sources from which women seek inductions, as recently demonstrated (Singh et al. 2017). In a prior small study, more than half of private abortion providers identified were not registered with the government to provide abortion, even though most were high quality providers (Ramachandar and Pelto 2004). In another larger assessment of abortion services in India, researchers determined that only 24% of all private abortion facilities in the country were certified and thereby legal (Duggal and Ramachandran 2004).

The official government statistics, and the responsiveness of reproductive health services that rely on these data, could also be greatly improved if more private abortion providers registered. To this end, it is important to improve the system for registering private clinics as official abortion providers, which currently can take months and even years to complete (Banerjee et al. 2015). A complementary strategy that would help to improve official government data on abortion services would be to provide additional Health Information Management Systems (HIMS) training related to the importance of logging induced abortion and PAC services.

Accurate estimates of induced abortion would underscore the need for public health campaigns emphasizing the legality of induced abortion and the dangers of seeking unsafe abortion providers or self-inducing. Studies have shown that anywhere from 31% to more than 50% of women who present at facilities with post-abortion complications have first attempted self-induction (Banerjee et al. 2012; Kumar et al.

2013). In the general population, one study found that 11% to 13% of women who reported experiencing an abortion had tried to self-induce, sought help from family or friends, or approached an untrained provider or chemist (Jejeebhoy et al. 2010). Most recently, researchers estimated that 11.5 million abortions, 73% of all annual abortions, were medical abortions done outside of health facilities (Singh et al. 2017). Information, education, and communication (IEC) campaigns in conjunction with provider trainings may increase the proportion of abortions performed in a safe, clinical setting with the appropriate guidance on dosage in the case of medical abortion.

Even with excellent research that makes compelling arguments for policy or programmatic change, the financial constraints and non-evidence based decision making that occurs in many policy environments is very real. As such, the policy implications of improved data on the social epidemiology of abortion, particularly unsafe abortion, are often limited by these and other factors. But while there is no guarantee that the desired policy changes will occur in the presence of strong research findings, changes are unlikely to occur in the absence of such results (Hardee et al. 2015). Thus, the overall objective of this research was to contribute to the larger dialogue around the public health implications of induced abortion occurring in India, ensuring that these important discussions continue to take place.

India legalized abortion more than 45 years ago, yet significant work remains in order to ensure equitable access to safe abortion for women throughout the country.

Policymakers must act to ensure adequate providers and services exist surrounding post-abortion and abortion care, as Indian women must control their own bodies if they are ever to achieve equality. Good abortion statistics would provide good data on disparities in access and motivate corrective policies. These data could thus help maximize the impact of limited resources by enabling creation of targeted, evidence-based interventions.

8. References

Anderson B, Katus K, Puur A, Silver B (1994) The Validity of Survey Responses on Abortion: Evidence from Estonia Demography 31:115-132

Arnold F, Kishor S, Roy TK (2002) Sex-Selective Abortions in India Population and Development Review 28:759-785

Aronow PM, Coppock A, Crawford FW, Green DP (2015) Combining list experiment and direct question estimates of sensitive behavior prevalence Journal of Survey Statistics and Methodology:smu023

Banerjee SK, Andersen KL, Navin D, Mathias G (2015) Expanding availability of safe abortion services through private sector accreditation: a case study of the Yukti Yojana program in Bihar, India Reproductive health 12:1

Banerjee SK, Andersen KL, Warvadekar J (2012) Pathways and consequences of unsafe abortion: A comparison among women with complications after induced and spontaneous abortions in Madhya Pradesh, India International Journal of Gynecology & Obstetrics 118:S113-S120

Begg MD, Parides MK (2003) Separation of individual-level and cluster-level covariate effects in regression analysis of correlated data Statistics in medicine 22:2591-2602 doi:10.1002/sim.1524

Biemer P, Brown G (2005) Model-based estimation of drug use prevalence using item count data Journal of Official Statistics 21:287

Blair G, Imai K (2012) Statistical Analysis of List Experiments Political Analysis 20:47-77 doi:10.1093/pan/mpr048

Blair G, Imai K, Lyall J (2014) Comparing and combining list and endorsement experiments: Evidence from Afghanistan American Journal of Political Science 58:1043-1063

Cannell CF, Miller PV, Oksenberg L (1981) Research on interviewing techniques Sociological methodology 12:389-437

Carpenter J, Bithell J (2000) Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians Statistcs in Medicine 19:1141-1164

Catania JA, Binson D, Canchola J, Pollack LM, Hauck W, Coates TJ (1996) Effects of interviewer gender, interviewer choice, and item wording on responses to questions concerning sexual behavior Public Opinion Quarterly 60:345-375

Center for Reproductive Rights (2013) The World's Abortion Laws Map 2013 Update. Center for Reproductive Rights, New York, NY

Comşa M, Postelnicu C (2013) Measuring social desirability effects on self-reported turnout using the item count technique International Journal of Public Opinion Research 25:153-172

Corstange D (2008) Sensitive Questions, Truthful Answers? Modeling the List Experiment with LISTIT Political Analysis 17:45-63 doi:10.1093/pan/mpn013

Coutts E, Jann B (2011) Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT) Sociological Methods & Research 40:169-193 doi:10.1177/0049124110390768

Cowan SK, Wu LW, Makela S, England P (2016) Alternative Estimates of Lifetime Prevalence Of Abortion from Indirect Survey Questioning Methods Perspectives on sexual and reproductive health

Davis RE, Couper MP, Janz NK, Caldwell CH, Resnicow K (2010) Interviewer effects in public health surveys Health education research 25:14-26 doi:10.1093/her/cyp046

Droitcour J, Caspar RA, Hubbard ML, Parsley TL, Visscher W, Ezzati TM (2004) The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application. In: Measurement Errors in Surveys. John Wiley & Sons, Inc., pp 185-210. doi:10.1002/9781118150382.ch11

Duggal R, Ramachandran V (2004) The Abortion Assessment Project–India: Key Findings and Recommendations Reproductive Health Matters 12:122-129 doi:http://dx.doi.org/10.1016/S0968-8080(04)24009-5

Edmeades J, Nyblase L, Malhotra A, MacQuarrie K, Parasuraman S, Walia S (2010) Methodological Innovations in Studying Abortion in Developing Countries: A "Narrative" Quantitative Survey in Madhya Pradesh, India Journal of Mixed Methods Research 4:176-198

Efron B (1987) Better bootstrap confidence intervals Journal of the American statistical Association 82:171-185

Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. CRC press,

Elul B (2004) Anonymous third party reporting of induced abortion: An experiment in Rajasthan, India. Paper presented at the Population Association of America, Boston, April 1-3

Elul B, Barge S, Verma S, Kumar N, Bracken H, Sadhvani H (2004) Unwanted pregnancy and induced abortion: data from men and women in Rajasthan, India. Citeseer,

Fernandez MM, Coeytaux F, de León RGP, Harrison DL (2009) Assessing the global availability of misoprostol International Journal of Gynecology & Obstetrics 105:180-186

Finer LB, Zolna MR (2016) Declines in unintended pregnancy in the United States, 2008–2011 New England Journal of Medicine 374:843-852

Foster DG Keynote Speaker Presentation on March 30th, 2016. In: Psychosocial Workshop, Washington, DC, 2016.

Foster DG (2017) Dramatic Decreases in US Abortion Rates: Public Health Achievement or Failure? American Public Health Association,

Fowler Jr FJ (2013) Survey research methods. Sage publications,

Fuentes L (2017) Personal communication regarding best friend method results compared to direct question in US survey.

Ganatra B et al. (2017) Global, regional, and subregional classification of abortions by safety, 2010–14: estimates from a Bayesian hierarchical model The Lancet

Glynn AN (2010) What Can We Learn with Statistical Truth Serum? Design and Alaysis of the List Experiment.

Glynn AN (2013) What Can We Learn with Statistical Truth Serum?: Design and Analysis of the List Experiment Public Opinion Quarterly 77:159-172 doi:10.1093/poq/nfs070

Gonzalez-Ocantos E, De Jonge CK, Meléndez C, Osorio J, Nickerson DW (2012) Vote buying and social desirability bias: Experimental evidence from Nicaragua American Journal of Political Science 56:202-217

Grice HP (1975) Logic and conversation. Syntax and semantics: 3 Speech acts. Academic Press, New York, New York

Grossman D et al. (2015a) Knowledge, opinion and experience related to abortion self-induction in Texas Contraception 92

Grossman D et al. (2015b) Knowledge, opinion and experience related to abortion self-induction in Texas Contraception 92:360-361

Grossman D, White K, Hopkins K, Potter JE (2014) The public health threat of anti-abortion legislation Contraception 89:73

Guttmacher Institute (2010) Methodologies for Estimating Abortion Incidence and Abortion-Related Morbidity: A Review. Guttmacher Institute, International Union for the Scientific Study of Population,

Hardee K, Wright K, Spicehandler J (2015) What is the Evidence That Evidence is Used in Family Planning Policy, Program and Practice Decision-Making? Paper presented at the Population Association of america, San Diego, CA,

Hawes M et al. (2017) Response patterns on behavioral outcomes in relation to use of resident enumerators over multiple survey rounds. Performance Monitoring and Accountability 2020 (PMA2020) Methodological Report No. 1. Bill & Melinda Gates Institute for Population and Reproductive Health, Johns Hopkins University, Baltimore, Maryland, USA

Holden RR, Kroner DG, Fekken GC, Popham SM (1992) A model of personality test item response dissimulation Journal of Personality and Social Psychology 63:272

Holtgraves T (2004) Social desirability and self-reports: Testing models of socially desirable responding Personality and Social Psychology Bulletin 30:161-172

Holtgraves T, Eck J, Lasky B (1997) Face Management, Question Wording, and Social Desirability1 Journal of Applied Social Psychology 27:1650-1671

Huntington D, Mensch B, Miller VC (1996) Survey questions for the measurement of induced abortion Stud Fam Plann 27:155-161

Imai K (2011) Multivariate Regression Analysis for the Item Count Technique Journal of American Statistical Association 106:407-416 doi:10.1198/jasa.2011.ap10415

International Institute for Population Sciences (IIPS) and Macro International (2007, ICF International [Distributor], 2007) National Family Health Survey (NFHS-3), 2005-06: India: Volume I [Dataset]. IAIR52FL.DTA. IIPS [Produders], Calverton, Maryland.

International Institute for Population Sciences (IIPS) and Macro International (2015-2016) National Survey of Family Health-4, 2015-2016: State Fact Sheets. International Institute for Population Sciences (IIPS), Mumbai, India

Jejeebhoy SJ (2011) Increasing access to safe abortion in rural Rajasthan: Outcomes of a Comprehensive Abortion Care Model. Population Council,

Jejeebhoy SJ, Kalyanwala S, Zavier AF, Kumar R, Jha N (2010) Experience seeking abortion among unmarried young women in Bihar and Jharkhand, India: delays and disadvantages Reproductive Health Matters 18:163-174

Joffe C (2009) Abortion and medicine: A sociopolitical history. In: Paul M, Lichtenberg E, Borgatta L, Grimes DA, Stubblefield P, Creinin M (eds) Management of Unintended and Abnormal Pregnancy. 1st edn. Wiley-Blackwell Publishing, New York, NY,

Jones EF, Forrest JD (1992) Contraceptive failure rates based on the 1988 NSFG Fam Plann Perspect 24:12-19

Jones RK, Jerman J (2017) Population group abortion rates and lifetime incidence of abortion: United States, 2008–2014 American journal of public health 107:1904-1909

Jones RK, Kost K (2007) Underreporting of Induced and Spontaneous Abortion in the United States: An Analysis of the 2002 National Survey of Family Growth Studies in Family Planning 38:187-197

Kanstrup C, Mäkelä M, Hauskov Graungaard A (2017) Women's reasons for choosing abortion method: A systematic literature review Scandinavian Journal of Public Health:1403494817717555

Kreuter F, Couper M, Lyberg L The use of paradata to monitor and manage survey data collection. In: Proceedings of the Joint Statistical Meetings, American Statistical Association, 2010. American Statistical Association Alexandria, VA, pp 282-296

Krosnick JA (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys Applied cognitive psychology 5:213-236

Kuklinski JH, Cobb MD, Gilens M (1997) Racial Attitudes and the "New South" The Journal of Politics 59:323-349

Kumar R, Zavier AF, Kalyanwala S, Jejeebhoy SJ (2013) Unsuccessful prior attempts to terminate pregnancy among women seeking first trimester abortion at registered facilities in bihar and jharkhand, india Journal of biosocial science 45:205-215

Langhaug LF, Cheung YB, Pascoe SJ, Chirawu P, Woelk G, Hayes RJ, Cowan FM (2011) How you ask really matters: randomised comparison of four sexual behaviour questionnaire delivery modes in Zimbabwean youth Sexually transmitted infections 87:165-173 doi:10.1136/sti.2009.037374

Lara D, Strickler J, Olavarrieta CD, Ellertson C (2004) Measuring Induced Abortion in Mexico Sociological Methods & Research 32:529-558 doi:10.1177/0049124103262685

Lensvelt-Mulders GJ, Hox JJ, Van der Heijden PG, Maas CJ (2005) Meta-analysis of randomized response research thirty-five years of validation Sociological Methods & Research 33:319-348

Lifewire (2017) What is GPRS? General Packet Radio Service. Lifewire. https://www.lifewire.com/general-packet-radio-service-817466. Accessed February 5 2017

Miller J (1984) A New Survey Technique for Studying Deviant Behavior. Ph.D Dissertation, The George Washington University

Ministry of Health and Family Welfare (2013) Health and Family Welfare Statistics of India 2013. Ministry of Health and Family Welfare, Statistics Division, New Delhi

Moseson H et al. (2015) Reducing under-reporting of stigmatized health events using the List Experiment: results from a randomized, population-based study of abortion in Liberia Int J Epidemiol doi:10.1093/ije/dyv174

Moseson H, Treleaven E, Gerdts C, Diamond-Smith N (2017a) The List Experiment for Measuring Abortion: What We Know and What We Need Studies in Family Planning doi:10.1111/sifp.12042

Moseson HS et al. Measuring Texas women's experiences with abortion self-induction using a list experiment. In: Contraception, 2017b. vol 4. Elsevier, p 272. doi:10.1016/j.contraception.2017.07.042

Nowakowska M (1970) A model of answering to a questionnaire item Acta Psychologica 34:420-439

Office of the Registrar General & Census Commissioner, India Ministry of Home Affairs, Government of India (2013) Annual Health Survey 2012-2013 Fact Sheet. Office of the Registrar General & Census Commissioner, New Delhi, India

Paulhus DL (1984) Two-component models of socially desirable responding Journal of personality and social psychology 46:598

Performance Monitoring and Accountability 2020 (2017) Performance Monitoring and Accountability 2020 (PMA2020). http://www.pma2020.org/sampling. Accessed April 5th 2017

Performance Monitoring and Accountability 2020 (PMA2020). http://www.pma2020.org/publications. Accessed November 12th 2015

Plummer ML, Wamoyi J, Nyalali K, Mshana G, Shigongo ZS, Ross DA, Wight D (2008) Aborting and Suspending Pregnancy in Rural Tanzania: An Ethnography of Young People's Beliefs and Practices Studies in Family Planning 39:281-292

R Development Core Team (2015) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Rahman M, DaVanzo J, Razzaque A (2014) Pregnancy termination in Matlab, Bangladesh: maternal mortality risks associated with menstrual regulation and abortion International perspectives on sexual and reproductive health 40:108-118

Ramachandar L, Pelto PJ (2004) Abortion providers and safety of abortion: a community-based study in a rural district of Tamil Nadu, India Reproductive Health Matters 12:138-146

Rodriguez LA, Sana M, Sisk B (2015) Self-administered questions and interviewer–respondent familiarity Field Methods 27:163-181

Rosenfeld B, Imai K, Shapiro J (2014) An empirical validation study of popular survey methodologies for sensitive questions Princeton University

Rossier C (2003) Estimating induced abortion rates: a review Stud Fam Plann 34:87-102

Rossier C, Guiella G, Ouedraogo A, Thieba B (2006) Estimating clandestine abortion with the confidants method--results from Ouagadougou, Burkina Faso Social science & medicine (1982) 62:254-266 doi:10.1016/j.socscimed.2005.05.024

Safi S et al. (2017) Resident Interviewers and Repeat Surveys: Effects on Measures of Reproductive Health. Paper presented at the Population Association of America, Chicago, IL, April 27-29, 2017

Sana M, Stecklov G, Weinreb AA (2016) A test of the stranger-interviewer norm in the Dominican Republic Population studies 70:73-92

Sedgh G et al. (2016) Abortion incidence between 1990 and 2014: global, regional, and subregional levels and trends The Lancet

Sedgh G, Singh S, Shah IH, Ahman E, Henshaw SK, Bankole A (2012) Induced abortion: incidence and trends worldwide from 1995 to 2008 Lancet 379:625-632 doi:10.1016/s0140-6736(11)61786-8

Simonds W, Ellertson C, Springer K, Winikoff B (1998) Abortion, revised: participants in the U.S. clinical trials evaluate mifepristone Social science & medicine (1982) 46:1313-1323

Singh S, Sedgh G, Hussain R (2010) Unintended Pregnancy: Worldwide Levels, Trends, and Outcomes Studies in Family Planning 41:241-250

Singh S et al. (2017) The incidence of abortion and unintended pregnancy in India, 2015 The Lancet Global Health 6:e111-e120 doi:10.1016/S2214-109X(17)30453-9

Singh S, Wittenberg J (2017) Personal communication from Susheela Singh and Jonathan Wittenberg at the Guttmacher Institute.

Sniderman PM, Carmines EG (1997) Reaching beyond Race PS: Political Science and Politics 30:466-471 doi:10.2307/420124

StataCorp (2015a) Stata 14 Base Reference Manual. Stata Press, College Station, TX

StataCorp (2015b) Statistical Software: Release 14. College Station, TX Patent,

StataCorp (2017) Stata 15 Base Reference Manual. Stata Press, College Station, TX

Stecklov G, Weinreb AA, Sana M (2015) Family Planning for Strangers: An Experiment on the Validity of Reported Contraceptive Use PloS one 10:e0136972 doi:10.1371/journal.pone.0136972

Stillman M, Frost JJ, Singh S, Moore AM, Kalyanwala S (2014) Abortion in India: a literature review New York: Guttmacher Institute

Sudman S, Bradburn NM (1974) Response effects in surveys. Chicago: Aldine,

Sudman S, Bradburn NM, Schwarz N (1996) Thinking about answers: The application of cognitive processes to survey methodology. Jossey-Bass,

Sundaram A, Vaughan B, Kost K, Bankole A, Finer L, Singh S, Trussell J (2017) Contraceptive Failure in the United States: Estimates from the 2006-2010 National Survey of Family Growth Perspectives on sexual and reproductive health 49:7-16 doi:10.1363/psrh.12017

Taylor SE (1981) The interface of cognitive and social psychology Cognition, social behavior, and the environment:189-211

Tourangeau R, Yan T (2007) Sensitive questions in surveys Psychological bulletin 133:859

Treleaven E, Thuy LD, Pham TN, Diamond-Smith N (2017) The List Experiment: Piloting a Methodology to Measure Stigmatized Behaviors around Sex-Selective Abortion in Vietnam. Paper presented at the IUSSP, Cape Town, South Africa,

Tsuchiya T, Hirai Y, Ono S (2007) A Study of the Properties of the Item Count Technique Public Opinion Quarterly 71:253-272 doi:10.1093/poq/nfm012

Weinreb AA (2006) The limitations of stranger-interviewers in rural Kenya American Sociological Review 71:1014-1039

Wolter F, Laier B The effectiveness of the item count technique in eliciting valid answers to sensitive questions. An evaluation in the context of self-reported delinquency. In: Survey Research Methods, 2014. vol 3. Citeseer, pp 153-168

Yeatman S, Trinitapoli J (2011) Best-friend reports: A tool for measuring the prevalence of sensitive behaviors American journal of public health 101:1666-1667

Zimmerman L, OlaOlorun F, Radloff S (2015) Accelerating and improving survey implementation with mobile technology: Lessons from PMA2020 implementation in Lagos, Nigeria 2015 29 doi:10.11564/29-1-719
1699-1712

Zimmerman L, Olson H, Tsui A, Radloff S (2017) PMA2020: Rapid Turn-Around Survey Data to Monitor Family Planning Service and Practice in Ten Countries Studies in Family Planning

SUZANNE O'DEA BELL, MPH

600 Acker Place NE
Washington, D.C. 20002
(916) 803-5249
suzanneobell@gmail.com

EDUCATION

2014 – Present	**Doctor of Philosophy (PhD) Candidate**, Department of Population Family and Reproductive Health, Demography Concentration Johns Hopkins University, Baltimore, MD
2009 – 2011	**Master of Public Health (MPH)**, Department of Maternal and Child Health, Global Public Health Focus University of California Berkeley, Berkeley, CA
2003 – 2007	**Bachelor of Arts (BA)**, Department of Communication Studies, Mass Communication Concentration, graduated *Magna Cum Laude* University of California Los Angeles, Los Angeles, CA

POSITIONS HELD

February 2016 – Present	**Graduate Student Research Assistant,** Dr. Caroline Moreau, Department of Population Family and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
August 2014 – Present	**Graduate Student Research Assistant**, Performance Monitoring and Accountability 2020, Gates Institute for Population and Reproductive Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
May 2017 – August 2017	**Contract Researcher**, Dr. Caitlin Gerdts, Ibis Reproductive Health, Boston, MA
August 2011 – July 2014	**Research Associate Specialist**, Bixby Center for Population Health and Sustainability, University of California Berkeley, Berkeley, CA

August 2010 – June 2011	**Graduate Student Researcher**, Bixby Center for Population Health and Sustainability, University of California Berkeley, Berkeley, CA
May 2010 – August 2012	**Research Fellow**, Misoprostol Pilot Project, Bixby Center for Population Health and Sustainability, University of California Berkeley, Lalmonirhat, Bangladesh
January – December 2008	**Volunteer Coordinator**, Support for International Change, Palo Alto, CA and Arusha, Tanzania

HONORS AND AWARDS

2017 – 2018	Laurie Schwab Zabin Award
2017 – 2018	Edward J. Dehne Award in Population Dynamics
2015 – 2017	Endowed Fellowship in Family Planning and Reproductive Health
2015	Gates Institute Travel Award
2015	Johns Hopkins Bloomberg School of Public Health Center for Global Public Health Research Scholarship
2014 – 2016	Robertson Scholarship
2010	University of California Berkeley Bixby Center for Population Health and Sustainability Practicum Fellowship
2009 – 2011	United States Department of Education Foreign Language Area Studies (FLAS) Fellowship (Swahili)

RESEARCH SUPPORT AND PROJECTS

Active

JHU Lead, “PMA2020 Multi-Country Study of Abortion”, Anonymous Donor. November 2017 – January 2019. (PI: S. Radloff)

JHU Lead, “Estimating the Incidence of Induced Abortion in Ghana”, DFID. October 2017 – September 2019. (PI: S. Keogh, C. Polis)

JHU Lead, “An In-Depth Assessment of the Impact of the Global Gag Rule and US Funding Cuts on Sexual and Reproductive Health Services and Outcomes in Sub-Saharan Africa”, Anonymous Donor. December 2017 – November 2018. (PI: G. Sedgh and E. Sully)

Graduate Research Assistant, “Prevalence of Medical Contraindications to Progestin-Only and Combined Oral Contraceptives in a Large Insurance Claims Database”, Society for Family Planning. August 2017 – March 2018. (PI: A. Burke)

Graduate Research Assistant, “Use of Claims and Electronic Medical Records to Assess Trends and Correlates of Quality Measures for Contraception Care", Bayer HealthCare Pharmaceuticals Inc., USA. November 2016 – March 2018. (PI: C. Moreau)

Graduate Research Assistant, "Bringing More Evidence into Contraceptive Counseling: Building a New Provider Tool to Tailor Predictions of Contraceptive Outcomes to Patient Sub-Populations", Society for Family Planning. October 2014 – March 2018. (PI: C. Moreau)

Graduate Research Assistant, "Performance Monitoring and Accountability (PMA) 2020", Bill and Melinda Gates Foundation. April 2013 – September 2018. (PI: S. Radloff)

Completed

Research Associate, "Scaling up Community-Based Distribution (CBD) of Depot Medroxyprogesterone Acetate (DMPA) in Ethiopia", Joffe Charitable Trust. September 2011 – February 2015. (PI: N. Prata)

Research Associate, "Preventing Postpartum Hemorrhage in Home Births in Rural Bangladesh", Venture Strategies for Health and Development. January 2009 – June 2011. (PI: N. Prata)

PEER REVIEWED PUBLICATIONS

Bell S, Zimmerman L, Choi Y, Hindin M (2017). Legal but Limited? Abortion Service Availability and Readiness Assessment in Nepal. Health Policy and Planning, doi: 10.1093/heapol/czx149.

Bell S, Bishai D (2017). Unmet need and sex: Investigating the role of coital frequency in fertility control. Studies in Family Planning, doi:10.1111/sifp.12012.

Weidert K, Gessessew A, **Bell S**, Godefay H, Prata N (2017). Community Health Workers as Social Marketers of Injectable Contraceptives: A Case Study from Ethiopia. Global Health: Science and Practice, https://doi.org/10.9745/GHSP-D-16-00344.

Nieto-Andrade B, Fidel E, Simmons R, Sievers D, Fedorova A, **Bell S**, Weidert K, Prata N (2017) Women's Limited Choice and Availability of Modern Contraception at Retail Outlets and Public-Sector Facilities in Luanda, Angola, 2012–2015. Global Health: Science and Practice, doi:10.9745/ghsp-d-16-00304.

Prata N, Downing J, **Bell S**, Weidert K, Godefay H, Gessessew A (2016). Cost of providing injectable contraceptives through a community-based social marketing program in Tigray, Ethiopia. Contraception, 93(6);485-491.

Prata N, **Bell S**, Weidert K, Nieto-Andrade B, Carvalho A, Neves I (2016). Varying family planning strategies across age categories: differences in factors associated with current modern contraceptive use among youth and adult women in Luanda, Angola. Open Access Journal of Contraception, doi: http://dx.doi.org/10.2147/OAJC.S93794.

Prata N, **Bell S**, Quaiyum MA (2014). Modeling maternal mortality in Bangladesh: the role of misoprostol in postpartum hemorrhage prevention. BMC Pregnancy Childbirth, 14:78.

Prata N, **Bell S**, Holston M, Quaiyum MA (2014). Is attendant at delivery associated with the use of interventions to prevent postpartum hemorrhage at home births? The case of Bangladesh. BMC Pregnancy Childbirth, 14:24.

Bell S, Passano P, Islam A, Prata N (2014). Training traditional birth attendants to prevent postpartum hemorrhage: A qualitative assessment. Journal of Health Population & Nutrition, 32(1);118-129.

Prata N, **Bell S**, Weidert K (2013). Prevention of postpartum hemorrhage in low-resource settings: current perspectives. International Journal of Women's Health, 5: 737-752.

Prata N, **Bell S**, Gessessew A (2013). Comprehensive abortion care: Evidence of improvements in hospital level indicators in Tigray, Ethiopia. BMJ Open, 3(7): e002873.

Prata N, **Bell S**, Weidert K, Gessessew A (2013). Potential for cost recovery: Women's willingness to pay for injectable contraceptives in Tigray, Ethiopia. PLoS ONE, 8(5): e64032.

Bell S, Prata N, Lahiff M, Eskenazi B (2012). Civil unrest and birthweight: An exploratory analysis of the 2007/2008 Kenyan Crisis. Social Science & Medicine, 74;1324-1330.

Prata N, Passano P, **Bell S**, Rowen T (2012). New hope: community-based misoprostol use to prevent postpartum hemorrhage. Health Policy and Planning, 28(4);339-346.

Prata N, Quaiyum MA, Passano P, **Bell S**, Bohl D, Hossain A, Azmi A, Begum M (2012). Training traditional birth attendants to use misoprostol and an absorbent delivery mat at home births. Social Science & Medicine, 75;2021-2027.

Prata N, Passano P, Rowen T, **Bell S** (2011). Where there are (few) skilled birth attendants. Journal of Health Population & Nutrition, 29(2);81-91.

Prata N, Hamza S, **Bell S**, Karasek D, Vahidnia F, Holston M. (2011) Inability to predict postpartum hemorrhage: insights from Egyptian intervention data. BMC Pregnancy & Childbirth, 11:97.

Prata N, **Bell S**, Holston M, Gerdts C, Melkamu Y (2011). Factors Associated with Choice of Post-Abortion Contraception in Addis Ababa, Ethiopia. African Journal of Reproductive Health, 15(3);55-62.

MANUSCRIPTS IN PROGRESS

Zimmerman L, **Bell S**, et al. Individual, neighborhood, and service delivery factors associated with modern contraceptive use: A multi-country analysis (*Under review*).

Moreau C, Gibbs S, **Bell S**, et al. Assessing family planning quality metrics using insurance claims data: Analysis from the Maryland all payer's claims database.

Bell S, Bishai D. Can a list experiment improve validity of abortion measurement?

Bell S, Bishai D. Paradata as a lens to understand underreporting of abortion at the individual level.

Bell S, Bishai D. The role of familiarity in reporting of abortion on surveys.

Alfonso N, **Bell S**, Bishai J, Chen M, Pan P, Shipley C, Bishai D. Impact of the Quality of Family Planning Services on Modern Contraceptive Prevalence.

Bishai D, Chen M, Pan Z, Alfonso N, **Bell S**, Shipley C. Where and for Whom Does Investing in Private-Sector Family Planning Quality Improve Contraceptive Prevalence?

Bell S, Safi S, Shankar M, Moreau C. Timing of contraceptive initiation with regard to first birth: A multi-country analysis.

Moore A, Browne A, **Bell S**. Capturing Medical Methods of Abortion Sales Data in India.

Sully E, Riley T, **Bell S**. Assessing Uncertainty in Expert Surveys: A Modified Approach to Measuring Abortion Incidence.

PEER REVIEWED ORAL PRESENTATIONS

Alfonso N*, **Bell S**, Bishai J, Chen M, Pan P, Shipley C, Bishai D. Impact of the Quality of Family Planning Services on Modern Contraceptive Prevalence. Population Association of America (PAA), Denver, 2018 (*Accepted*).

Bishai D*, Chen M, Pan Z, Alfonso N, **Bell S**, Shipley C. Where and for Whom Does Investing in Private-Sector Family Planning Quality Improve Contraceptive Prevalence? Population Association of America (PAA), Denver, 2018 (*Accepted*).

Bell S*, Safi S, Shankar M, Moreau C. Timing of first contraceptive use with regard to first birth: A multi-country analysis. Population Association of America (PAA), Chicago, 2017.

Bell S*, Bishai D. The sex dividend: The association between unmet need and coital frequency. International Conference on Family Planning, Nusa Dua, Indonesia, 2016.

Prata N, **Bell S***, Weidert K, Nieto-Andrade B, Carvalho A, Neves I (2016). Varying family planning strategies across age categories: differences in factors associated with current modern contraceptive use among youth and adult women in Launda, Angola. International Conference on Family Planning, Nusa Dua, Indonesia, 2016.

Bell S*, Prata N, Lahiff M, Eskenazi B. Civil unrest and birthweight: An exploratory analysis of the 2007/2008 Kenyan Crisis. Population Association of America (PAA), San Francisco, CA, 2012.

PEER REVIEWED POSTER PRESENTATIONS

Bell S, Bishai D. Paradata as a lens to understand underreporting of abortion at the individual level. Population Association of America (PAA), Denver, 2018 (*Accepted*).

Sully E*, Riley T, **Bell S**. Philbin J. Assessing Uncertainty in Expert Surveys: A Modified Approach to Measuring Abortion Incidence. Population Association of America (PAA), Denver, 2018 (*Accepted*).

Moore A*, Browne A, **Bell S**. Capturing Medical Methods of Abortion Sales Data in India. Population Association of America (PAA), Denver, 2018 (*Accepted*).

Bell S*, Bishai D. The Sex Dividend: The Association Between Unmet Need and Coital Frequency. Population Association of America (PAA), Washington DC, 2016.

Prata N, **Bell S***, Weidert K, (2013). Potential for Cost Recovery: Women's Willingness to Pay for Injectable Contraceptives in Tigray, Ethiopia. Population Association of America (PAA), New Orleans, 2013.

INVITED PRESENTATIONS

Bell S, "The list experiment in Rajasthan, India". Expert Meeting on Abortion Measurement Methodologies, Guttmacher Institute, New York, NY, September 25th, 2017.

TEACHING EXPERIENCE

Fall 2017	**Graduate Teaching Assistant**, 380.600: Principles of Population Change (Graduate Level), Johns Hopkins Bloomberg School of Public Health
Spring 2016, 2017	**Graduate Teaching Assistant and Lecturer**, 380.840: Public Health Perspectives on Abortion (Graduate Level), Johns Hopkins Bloomberg School of Public Health

August 2015	**Invited Lecturer**, 5-Day Workshop on Data Analysis in Stata: Centre for Research, Evaluation Resources and Development, Ile-Ife, Nigeria
Fall 2012, 2013	**Invited Speaker**, PB HLTH 293: Maternal and Child Health Department Master's Thesis Seminar (Graduate Level), University of California Berkeley

SERVICE AND COMMITTEE LEADERSHIP

2016 – Present	Co-Coordinator, Public Health Students for Choice, Johns Hopkins Bloomberg School of Public Health
2016	Proposal Reviewer, Packard Foundation International Conference on Family Planning Grant Competition
2015	Abstract Reviewer, International Conference on Family Planning – Indonesia
2015 – 2016	President, Public Health Students for Choice, Johns Hopkins Bloomberg School of Public Health
2015 – 2016	President, Abortion Journal Club, Johns Hopkins Bloomberg School of Public Health
2010 – 2011	Student Representative, Department of Maternal and Child Health Admissions Committee, University of California Berkeley
Summer 2007	Volunteer, Support for International Change, Arusha, Tanzania

PROFESSIONAL ASSOCIATIONS

2018 – Present	Member, Society for Family Planning
2017 – Present	Member, International Union for the Scientific Study of Population
2012, 2016 – Present	Member, Population Association of America
2015 – Present	Member, National Abortion Federation